

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

QM nucleic - nucleic search, using sw model  
Run on: January 31, 2004, 09:06:13 ; Search time 7412 seconds  
(without alignments)  
10994.600 Million cell updates/sec

Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: 1 gtaggaactctcaaaagt.....tttaataataaaaaaaa 1992

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapex 1.0  
Searched: 288711-segs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:

- 1: gb\_ba:
- 2: gb\_htg:
- 3: gb\_in:
- 4: gb\_ov:
- 5: gb\_ov:
- 6: gb\_pat:
- 7: gb\_ph:
- 8: gb\_pi:
- 9: gb\_pi:
- 10: gb\_ro:
- 11: gb\_ro:
- 12: gb\_sy:
- 13: gb\_un:
- 14: gb\_vi:
- 15: em\_ba:
- 16: em\_fun:
- 17: em\_hum:
- 18: em\_in:
- 19: em\_mu:
- 20: em\_or:
- 21: em\_or:
- 22: em\_ov:
- 23: em\_pat:
- 24: em\_ph:
- 25: em\_pi:
- 26: em\_ro:
- 27: em\_ro:
- 28: em\_ro:
- 29: em\_ro:
- 30: em\_vt:
- 31: em\_vt:
- 32: em\_vt:
- 33: em\_vt:
- 34: em\_vt:
- 35: em\_vt:
- 36: em\_vt:
- 37: em\_vt:
- 38: em\_vt:
- 39: em\_vt:
- 40: em\_vt:
- 41: em\_vt:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1992	100.0	1992	6	AX054819 Sequence
2	1593	80.0	1593	9	AF121863 Homo sapi
3	1540.4	77.3	3038	9	BC005110 Homo sapi
4	1530	76.8	2925	9	BC046520 Homo sapi
5	1525.2	76.6	3616	9	AK095380 Homo sapi
6	1516.6	76.1	3576	9	AK026479 Homo sapi
7	1504	75.5	3145	6	AX512835 Sequence
8	1504	75.5	3145	6	AX000362 Homo sapi
9	1371.4	68.8	2661	9	AY044865 Homo sapi
10	1193.2	59.9	1782	10	BC043328 Mus muscu
11	925.4	46.5	968	9	AF420561 Homo sapi
12	516	25.9	129010	9	AL589666 Human DNA
13	427.2	21.4	451	6	AX331384 Sequence
14	425	21.3	425	6	BD113855 EST and e
15	357	17.9	358	6	BD026500 Sequence
16	172	8.6	278898	2	AC116713 Mus muscu
17	167.2	8.4	224551	2	AC111832 Rattus no
18	163.8	8.2	454	11	G30543 human STS S
19	135.6	6.8	278375	2	AC130093 Rattus no
20	86	4.3	311	6	AX185131 Sequence
21	86	4.3	313	6	AX186466 Sequence
22	84	4.2	299	6	AX188027 Sequence
23	78.6	3.9	580	6	AX387326 Sequence
24	75.4	3.8	167444	9	AP005014 Homo sapi
25	73.4	3.7	183972	2	EX537259 Danio rer
26	72.8	3.7	211118	2	EX284684 Danio rer
27	71.8	3.6	110000	2	PFMAL8P1_12
28	71.2	3.6	125623	3	AC115599 Dictyoste
29	70	3.5	156441	9	AC091869 Homo sapi
30	70	3.5	168885	9	AC091873 Homo sapi
31	69.6	3.5	333321	3	AC116986 Dictyoste
32	69.2	3.5	187310	2	EX005122 Danio rer
33	69	3.5	160411	2	AC135935 Rattus no
34	68.4	3.4	9095	6	AX458545 Sequence
35	68.4	3.4	14422	3	AF466146 Melipona
36	67.4	3.4	14012	3	AE001433 Plasmodiu
37	67.4	3.4	343050	3	PF929353 Plasmodiu
38	67.2	3.4	8056	6	AX599046 Sequence
39	67.2	3.4	195620	2	EX088600 Danio rer
40	67	3.4	141913	9	AL512452 Human DNA
41	66.8	3.4	348600	1	AB063521 Wigglewo
42	66.2	3.3	7461	6	AX346687 Sequence
43	65.8	3.3	161757	2	AC044901 Homo sapi
44	65.6	3.3	6161	6	AX345525 Sequence
45	65.6	3.3	8056	6	AX598900 Sequence

ALIGNMENTS

RESULT 1	AX054819	1992 bp	DNA	linear	FAT 13-JAN-2001
LOCUS	Sequence 3 from Patent WO0073334.				
DEFINITION	Sequence 3 from Patent WO0073334.				
ACCESSION	AX054819				
VERSION	AX054819.1	GI:12228268			
KEYWORDS	Homo sapiens (human)				
SOURCE	Homo sapiens				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
REFERENCE	1				
AUTHORS	Yue, H., Tang, Y.T. and Azimzai, Y.				
TITLE	Human sorting nexins				
JOURNAL	Patent: WO 0073334-A 3 07-DEC-2000;				

[illegible]



[illegible]



QY	713	GAC	TTTCTTTCCCTTAATGCTGGGGAACA	CAATTTCTTGATAAGATACTAC	CAGATGTA	772
Db	2049	GAC	TTTCTTTCCCTTAATGCTGGGGAACA	CAATTTCTTGATAAGATACTAC	CAGATGTA	2108
QY	773	AAT	CTTTGGGAAAAATTATAAATCTG	TCTCGGAAACTAATGAAAGAGAGT	CAGCAT	832
Db	2109	AAT	CTTTGGGAAAAATTATAAATCTG	TCTCGGAAACTAATGAAAGAGAGT	CAGCAT	2168
QY	833	TTG	GAACCTTTTATCATGAATTTCA	TAAATCTTTGTGAGTCTCAAAGCCT	AAACCAAGT	892
Db	2169	TTG	GAACCTTTTATCATGAATTTCA	TAAATCTTTGTGAGTCTCAAAGCCT	AAACCAAGT	2228
QY	893	AGAC	CGAAGCTGACCAATCTCAG	CCCTACTTCAGAAACAAACAAGAGCT	TTTTCAATGAT	952
Db	2229	AGAC	CGAAGCTGACCAATCTCAG	CCCTACTTCAGAAACAAACAAGAGCT	TTTTCAATGAT	2288
QY	953	CTG	TTTAAATAATATGCAAAACCG	TGCTGAAAAATACAGAGAGAAAC	AAATCAGATTAT	1012
Db	2289	CTG	TTTAAATAATATGCAAAACCG	TGCTGAAAAATACAGAGAGAAAC	AAATCAGATTAT	2348
QY	1013	TTT	ATGAGGTGATGACTGTAGAGGAG	TCATATGATTACCTGATGTATG	PAGACGGGTA	1072
Db	2349	TTT	ATGAGGTGATGACTGTAGAGGAG	TCATATGATTACCTGATGTATG	PAGACGGGTA	2408
QY	1073	GT	TTTCCAGGTCTCTGACTGGCT	TCATCATCTCTTAATGGGAACT	CGAATCTCTTTAAA	1132
Db	2409	GT	TTTCCAGGTCTCTGACTGGCT	TCATCATCTCTTAATGGGAACT	CGAATCTCTTTAAA	2468
QY	1133	AAC	ACCTGGAATATGATATCTGAT	TACTATCTTCCTCAGTGTAAC	ATAGTATTTTCAG	1192
Db	2469	AAC	ACCTGGAATATGATATCTGAT	TACTATCTTCCTCAGTGTAAC	ATAGTATTTTCAG	2528
QY	1193	GAG	CACCGTTTGGTCTCACTCAT	ATAACACTTCTCAGAGATGCT	ATATCTGTGAAACACCT	1252
Db	2529	GAG	CACCGTTTGGTCTCACTCAT	ATAACACTTCTCAGAGATGCT	ATATCTGTGAAACACCT	2588
QY	1253	GAAC	CTCGCTCTCTCCAAAGATAAG	CAAAAAAGGAGCAAAA	CAGACTTTTGAAGAAATGATG	1312
Db	2589	GAAC	CTCGCTCTCTCCAAAGATAAG	CAAAAAAGGAGCAAAA	CAGACTTTTGAAGAAATGATG	2648
QY	1313	AAT	TACATTCAGATCTGTTAGT	CAAGTATTTGGTGAAGAAAC	CAAGTATGAAGCATC	1372
Db	2649	AAT	TACATTCAGATCTGTTAGT	CAAGTATTTGGTGAAGAAAC	CAAGTATGAAGCATC	2708
QY	1373	AGAC	TTCTGTTTGGTGTACAGCAAC	CAGTACTCAACAGCAGCTGAC	TATGTTTTA	1432
Db	2709	AGAC	TTCTGTTTGGTGTACAGCAAC	CAGTACTCAACAGCAGCTGAC	TATGTTTTA	2768
QY	1433	TTG	GACATTTGTGATACAGGAAC	TGTTTCCAGAGCTCAAT	AAGGTACAAAAGGAAGTTACC	1492
Db	2769	TTG	GACATTTGTGATACAGGAAC	TGTTTCCAGAGCTCAAT	AAGGTACAAAAGGAAGTTACC	2828
QY	1493	TCT	GTGACATCTTGGATGTAAAC	CTCTGGATTTGGTATAGAA	TAAACCATTTGAAATTTCT	1552
Db	2829	TCT	GTGACATCTTGGATGTAAAC	CTCTGGATTTGGTATAGAA	TAAACCATTTGAAATTTCT	2888
QY	1553	GCT	GTGCGAGGCTGTAGAAATTT	CTTTTTTGGGTATATCT	TATATATATATATGTA	1612
Db	2889	GCT	GTGCGAGGCTGTAGAAATTT	CTTTTTTGGGTATATCT	TATATATATATATGTA	2948
QY	1613	TCG	CTGTCTGAAATTTTAGTTAT	TTTTTTTGTATTTTAA	TAAAGACTTAAACAACTTAATGA	1672
Db	2949	TCG	CTGTCTGAAATTTTAGTTAT	TTTTTTTGTATTTTAA	TAAAGACTTAAACAACTTAATGA	3008
QY	1673	TTAAA	AGTGA	1682		
Db	3009	TTAAA	AGTGA	3018		

RESULT 4	2925 bp	1 linear	mrna	PRI 21-FEB-2003
BC046520				
LOCUS				

DEFINITION	Homo sapiens, Similar to sorting nexin 14, clone IMAGE:5267454, mRNA.
ACCESSION	BC046520
VERSION	BC046520.1 GI:28461364
KEYWORDS	
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE	1 (bases 1 to 2925)
JOURNAL	Strausberg, R.
AUTHORS	Direct Submission
REMARK	Submitted (03-FEB-2003) National Institutes of Health, Mammalian
COMMENT	Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550, USA
	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
	Contact: MGC help desk
	Email: <a href="mailto:cgabs@mail.nih.gov">cgabs@mail.nih.gov</a>
	Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
	cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
	DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
	Center code: BCM-HGSC
	Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a>
	Contact: <a href="mailto:amg@bcm.tmc.edu">amg@bcm.tmc.edu</a>
	Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegod, H., Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati, A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAP Plate: 91 Row: a Column: 6.

## FEATURES

```

1. 2925
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cd_name="IMAGE:5267454"
/tissue_type="Testis"
/clone_lib="XIH MGC_97"
/lab_host="DH10B"
/note="vector; paluescri

```

BASE COUNT	ORIGIN
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

Query Match	76.8%	Score 1530	DB 9	Length 2925
Best Local Similarity	96.3%	Pred. No. 1.8e-280		
Matches 1566:	Conservative	0: Mismatches 60	Indels 0	Gaps 0

QY	53	AAACTGTTTCTCTTTTTTTTTTAAATAGTTTTTTCACATAAACTAGACTGTGTGTGATAATTT	112
Db	1288	AATGTATTTACTCCTATGTTCTGCCATAGTGAATGAGTATTTTCACACAACTTTTAAAGAGGT	1347
QY	113	GCTATGTACTTGTATACATATTTTGTGTTTAAATATTTCAGGAACACACAGAAAAGGGGAGANTCA	172
Db	1348	CGAATACACCAACACGCCAATTCAAAATTTGAAACAGGAACACACAGAAAAGGGGAGANTCA	1407
QY	173	TTTGGAAATCAGCAGAAATAGTGTACCAAAATTAAGGAGTATTTCAAAGTACCACAATCGAG	232
Db	1408	TTTGGAAATCAGCAGAAATAGTGTAGCAAAATTAAGGAGTATTTCAAAGTACCACAATCGAG	1467
QY	233	GGAGCTATGTTGCTCAATTATGTTGTAGCTGAAGGTCAAGATGATTTTATTGAAGAAGGT	292
Db	1468	GGAGCTATGTTGCTCAATTATGTTGTAGCTGAAGGTGAAGTATTTTATTGAAGAAGGT	1527
QY	293	ATTCTTCTAATGGAAGATGATTTCTCCAGTGGAGGCTGTGAGCACACCTAATACTCCCCGA	352
Db	1528	ATTGTTGTATGGAAGATGATTTCTCCAGTGGAGGCTGTGAGCACACCTAATACTCCCCGA	1587
QY	353	AACCTTGTGTCATGGAAAAATTAGCAATTCCATATGTAGACTTTTTTTTGGAGATCCCTCTCTCT	412

RESULT 4  
BC046520  
LOCUS

Db	2668	TTGGACATTGTGATACAGGAACCTGTTTCCAGAGCTCAATTAAGGTACAAAGGAAGTTACC	2722
Qy	1493	TCGTGACATCTTTGGATGTAAACACATCTGGATTGGTATAGAAATTAACCCATTGAAATTTCT	1552
Db	2728	TCGTGACATCTTTGGATGTAAACACATCTGGATTGGTATAGAAATTAACCCATTGAAATTTCT	2787
Qy	1553	GCTGTGCGAGGGTGTAGAAATTTACTTTTTTTGGGTATATTCCTTATATATATTTATGTACA	1612
Db	2788	GCTGTGCGAGGGTGTAGAAATTTACTTTTTTTGGGTATATTCCTTATATATATTTATGTACA	2847
Qy	1613	TCGCTGCTGTAATTTTAGTTATTTTTTTGTTTTTAAATAAGACTTAACACAACTTAATGA	1672
Db	2848	TCGCTGCTGTAATTTTAGTTATTTTTTTGTTTTTAAATAAGACTTAACACAACTTAATGA	2907
Qy	1673	TTAAAA 1678	
Db	2908	TTAAAA 2913	
RESULT 5			
AKO95380			
LOCUS			
DEFINITION	AKO95380	3616 bp mRNA linear	PRI 15-JUL-2002
		homo sapiens cDNA FLJ38061 fis, clone CTONG2014966, highly similar	
		to SORTING NEXIN 14.	
ACCESSION	AKO95380		
VERSION	AKO95380.1	GI:21754626	
KEYWORDS		oligo capping; fis (full insert sequence).	
SOURCE		Homo sapiens (human)	
ORGANISM		Homo sapiens	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
		1	
		Tanigami,A., Fujiwara,T., Shibahara,T., Goto,Y., Hirao,M., Shimizu,F., Wakebe,H., Ono,T., Hishigaki,H., Watanabe,T., Ozaki,K., Sugiyama,T., Irie R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Sugiyama,A., Kawakami,B., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.	
		NEDO human cDNA sequencing project	
TITLE		Unpublished	
JOURNAL		2 (bases 1 to 3616)	
REFERENCE		Isogai,T. and Yamamoto,J.	
AUTHORS		Direct Submission	
TITLE		Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7	
JOURNAL		Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan	
		(E-mail:genomics@kri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)	
COMMENT		NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.	
FEATURES		Location/Qualifiers	
source		1..3616	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="CTONG2014966"	
		/tissue_type="tongue, tumor tissue"	
		/clone_lib="CTONG2"	
		/note="Cloning vector: pME18SFL3"	
BASE COUNT	1078 a	629 c	710 g
ORIGIN		1199 t	
Query Match	76.6%	Score 1525.2;	DB 9; Length 3616;
Best Local Similarity	96.4%;	Pred. No. 1.2e-259;	
Matches 1571; Conservative	0; Mismatches	58; Indels	1; Gaps
Qy	53	AAACTGTTTTCTTTTTTAAATAGTGTTCCTTCAATAAATAGACTGTGTGTGATAATTT	112

1987 AATGATTTTACCTCTATGTTCTGCCATAGTATGATGATTTTCAGACAACTTTTAAAGGT 2046  
113 GCTATGTTACTTGATACATTTTGTGTTTAAATATTCAGGAAACACACAGAAAGGGGAGATCA 172  
2047 GCAGAAATCACCAACCGCAATTCAAAATTTGAACAGGAAACACACAGAAAGGGGAGATCA 2106  
173 TTTGGAAATCAGCAGAAATAGGTAGCAAAATTAAGGAGTATTCAAAAGTACCAATGGAG 232  
2107 TTTGGAAATCAGCAGAAATAGGTAGCAAAATTAAGGAGTATTCAAAAGTACCAATGGAG 2166  
233 GGAGCTATGCTGCTAAATATGTTGCTGAGCTGAGGTGAGATGATTTTAAATGAAGAAGGT 292  
2167 GGAGCTATGCTGCTAAATATGTTGCTGAGCTGAGGTGAGATGATTTTAAATGAAGAAGGT 2226  
293 ATTGTTGTAATGGAAGATGATTTCTCCAGTGGAGGCTGTGAGCACCACTAACTACCTCCCGA 352  
2227 ATTGTTGTAATGGAAGATGATTTCTCCAGTGGAGGCTGTGAGCACCACTAACTACCTCCCGA 2286  
353 AACCTTGTCTGCTGGAATATGAGATTCATATGATGAGATTTTTCAGGATCCCTCTCT 412  
2287 AACCTTGTCTGCTGGAATATGAGATTCATATGATGAGATTTTTCAGGATCCCTCTCT 2346  
413 GAAAGGAGGAG-AAAAAGGAGATTCCTGTTGTTTCTGATGATGTTGAAAGAAATGA 471  
2347 GAAAGGAGGAGAAAAAGGAGATTCCTGTTGTTTCTGATGATGTTGAAAGAAATGA 2406  
472 TAGAAGAGAGTGTGACACAGAGCTGAAATGTTGTTCTGTCTATAGAGATATCTTGAAT 531  
2407 TAGAAGAGAGTGTGACACAGAGCTGAAATGTTGTTCTGTCTATAGAGATATCTTGAAT 2466  
532 CTATGTTGTAATCAAAATCAACAGAAATTCATGTTGATGATTTCTGATGCCAGCTTCC 591  
2467 CTATGTTGTAATCAAAATCAACAGAAATTCATGTTGATGATTTCTGATGCCAGCTTCC 2526  
592 TTCTAGAGAGATCATTTGGCCCCCAAAATTTATGAAATTTCTAAAGTCAAGAGGAGAGATT 651  
2527 TTCTAGAGAGATCATTTGGCCCCCAAAATTTATGAAATTTCTTAAAGTCAAGAGGAGAGATT 2586  
652 CCAAGATATCTACAGAAATCTCTGAGCATCCAGAACTGAGTAATAGTCAACTCTGTC 711  
2587 CCAAGATATCTACAGAAATCTCTGAGCATCCAGAACTGAGTAATAGTCAACTCTGTC 2646  
712 AGACTTTCTTTCCCTTAATGTTGGGAAAACACAAATTTCTTGAATGATATACAGAGAT 771  
2647 AGACTTTCTTTCCCTTAATGTTGGGAAAACACAAATTTCTTGAATGATATACAGAGAT 2706  
772 AAATCTTTGGGAAAATTAATAAATCTGTTCTGGAATACTAATGAAGAGAGAGGTGAGCA 831  
2707 AAATCTTTGGGAAAATTAATAAATCTGTTCTGGAATACTAATGAAGAGAGAGGTGAGCA 2766  
832 TTTGGAACTTTTATCATGAAATTTTCAATTTCTTGTGAGTCTCCAAAGCCTAAACCAAG 891  
2767 TTTGGAACTTTTATCATGAAATTTTCAATTTCTTGTGAGTCTCCAAAGCCTAAACCAAG 2826  
892 TAGACCAAGACTGACCACTCTCAGCCCTACTTCAGAAAACACAAAGAGCTTTTCAATGA 951  
2827 TAGACCAAGACTGACCACTCTCAGCCCTACTTCAGAAAACACAAAGAGCTTTTCAATGA 2886  
952 TCTGTTTAAAATTAATCAACCGTGTGAAAATACAGAGAGAAAGCAAAATCAGAAATTA 1011  
2887 TCTGTTTAAAATTAATCAACCGTGTGAAAATACAGAGAGAAAGCAAAATCAGAAATTA 2946  
1012 TTTATGAGAGTGTAGCTGTAGAGAGTCTATGATTTACTGATGTATGTAGAGCGGT 1071  
2947 TTTATGAGAGTGTAGCTGTAGAGAGTCTATGATTTACTGATGTATGTAGAGCGGT 3006  
1072 AGTTTCCAGGTTCTGAGCTGCTCATCTCTTAAATGGAACCTGGAATCTCTTTTAA 1131  
3007 AGTTTCCAGGTTCTGAGCTGCTCATCTCTTAAATGGAACCTGGAATCTCTTTTAA 3066  
1132 AAACACCTCGAAATGATATCTGATTTACTATCTTCAAGTGTAAATAGAACAGCTATTTCA 1191

3067 AAACACCTCGAAATGATATCTGATTTATCTCAGTGTAAACTAGAACACGCTATTTCA 3126  
1192 GGAGCACCCTGTTGTTCTCACTCATTAACACTTCTCAGAGATGCTATATTTCTGTGAAAAACAC 1251  
3127 GGAGCACCCTGTTGTTCTCACTCATTAACACTTCTCAGAGATGCTATATTTCTGTGAAAAACAC 3186  
1252 TGAACCTCGCTCTCTCCAAAGATAAGCAAAAAGGAGCAAAACAGACTTTTGAAGAAATGAT 1311  
3187 TCAACCTCGCTCTCTCCAAAGATAAGCAAAAAGGAGCAAAACAGACTTTTGAAGAAATGAT 3246  
1312 GAATTAATTCAGATCTGTTAGTCAAGTGTATTTGTTGAGGAAACCAAGTATGAAGCAT 1371  
3247 GAATTAATTCAGATCTGTTAGTCAAGTGTATTTGTTGAGGAAACCAAGTATGAAGCAT 3306  
1372 CAGACTTCTGTTGTTGATGGCTTACAGCAACCAAGTCTCAACCAAGCAGCTGACTTATGTTT 1431  
3307 CAGACTTCTGTTGTTGATGGCTTACAGCAACCAAGTCTCAACCAAGCAGCTGACTTATGTTT 3366  
1432 ATTGGCAATTCAGATGATCAGGAACTGTTTCCAGAGCTCAATAAGSTACAAAAGGAGTTAC 1491  
3367 ATTGGCAATTCAGATGATCAGGAACTGTTTCCAGAGCTCAATAAGSTACAAAAGGAGTTAC 3426  
1492 CTCTGTGACATCTTGGATGTAAACACATTTGGATTTGGTATGAAATACCCATTCGAAATTC 1551  
3427 CTCTGTGACATCTTGGATGTAAACACATTTGGATTTGGTATGAAATACCCATTCGAAATTC 3486  
1552 TGCTGTGCGAGGTTGTTAGTAAATTTTCTTTTGGTATATTTCTTATATATATATATGAT 1611  
3487 TGCTGTGCGAGGTTGTTAGTAAATTTTCTTTTGGTATATTTCTTATATATATATGAT 3546  
1612 ATCGCTGCTGAAATTTAGTATTTTCTTTTAAATAAGACTTAACACAACTTAATG 1671  
3547 ATCGCTGCTGAAATTTAGTATTTTCTTTTAAATAAGACTTAACACAACTTAATG 3606  
1672 ATTAAGAGTG 1681  
3607 ATTAAGAGTG 3616

## RESULT 6

AK026479

LOCUS

DEFINITION

AK026479.1

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

3576 bp mRNA linear PRI 29-SEP-2000  
Homo sapiens cDNA: FLJ22826 fis, clone KALIA4022, highly similar to  
AF121863 Homo sapiens sorting nexin 14 (SNX14) mRNA.

AK026479.1 GI:10439350  
oligo capping; fis (full insert sequence).

Homo sapiens (human)

Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (sites)

Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T.,  
Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Tanigami, A.,  
Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hiro, M.,  
Ohmori, Y., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T.,  
Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.

NEDO human cDNA sequencing project

Unpublished

2 (bases 1 to 3576)

Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
Shibahara, T., Tanaka, T. and Nakamura, Y.

Direct Submission

Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail: cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)

NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,





AUTHORS Daemen,M.J., Cleutjens,C.B. and Zaman,G.J.  
 TITLE Markers of unstable atherosclerotic plaques  
 JOURNAL Patent: WO 02062839-A 4 15-AUG-2002;  
 Universiteit Maastricht (NL)  
 FEATURES Location/Qualifiers  
 source 1..3145

/organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 BASE COUNT 985 a 570 c 659 g 931 t  
 ORIGIN

Query Match 75.5%; Score 1504; DB 6; Length 3145;  
 Best Local Similarity 99.5%; Pred. No. 7e-256;  
 Matches 1529; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

143 TTGAGAAACACAGAGAAAGGGAGAAATCATTTGGAATCAGCAGAAATAGGTAGCAAAATT 202  
 Db TTTCGAAACACAGAGAAAGGGAGAAATCATTTGGAATCAGCAGAAATAGGTAGCAAAATT 1657  
 QY 203 AAGAGGATTCMAAGTACCACATGGAGGAGCTATGTTGCCCTAATATATGGTGTAGCT 262  
 Db 1658 AAGAGGATTCAGAAATGACCAATGGAGGAGCTATGTTGCCCTAATATATGGTGTAGCT 1717  
 QY 263 GAAGGTGAAGATGATTTTATGAAGAGGTATTTGTTAAATGGAAGATGATTTCTCCAGTG 322  
 Db 1718 GAAGGTGAAGATGATTTTATGAAGAGGTATTTGTTAAATGGAAGATGATTTCTCCAGTG 1777  
 QY 323 GAGCGTGTGAGCACACCTAATATCTCCCGAAACCTTGTCTGATGGAATATAGCAATTTCCA 382  
 Db 1778 GAGCGTGTGAGCACACCTAATATCTCCCGAAACCTTGTCTGATGGAATATAGCAATTTCCA 1837  
 QY 383 TATGTAGATTTTGTGAGGATCTCTCTGAAAGAGGAGAAAGAAAGAAATTTCTT 442  
 Db 1838 TATGTAGATTTTGTGAGGATCTCTCTGAAAGAGGAGAAAGAAAGAAATTTCTT 1897  
 QY 443 GTGTTTGTATGATCTTGAAGAAATGATAGAGCAGAGTGGACAGGCGCTGAACAT 502  
 Db 1898 GTGTTTGTATGATCTTGAAGAAATGATAGAGCAGAGTGGACAGGCGCTGAACAT 1957  
 QY 503 TGGTCTGTCTATAGAGATATCTTGAATCTATGTTACTTGAATCAAACTCAACAGAAATT 562  
 Db 1958 TGGTCTGTCTATAGAGATATCTTGAATCTATGTTACTTGAATCAAACTCAACAGAAATT 2017  
 QY 563 CATGGTGCAATTTCTGATGCCAGCTTCTCTTAAAGAGATCATTTGGCCCCCAAAATTTAT 622  
 Db 2018 CATGGTGCAATTTCTGATGCCAGCTTCTCTTAAAGAGATCATTTGGCCCCCAAAATTTAT 2077  
 QY 623 GAATTTCTTAAAGTCAAGAGGGAGAGTTCCAGAAATATCTACAGAAATCTTCGAGCAT 682  
 Db 2078 GAATTTCTTAAAGTCAAGAGGGAGAGTTCCAGAAATATCTACAGAAATCTTCGAGCAT 2137  
 QY 683 CCAGAACTGATTAATAGTCAACTTCTGGCAGACTTTCTTCCCTAATGTTGGGGAACA 742  
 Db 2138 CCAGAACTGATTAATAGTCAACTTCTGGCAGACTTTCTTCCCTAATGTTGGGGAACA 2197  
 QY 743 CAATTTCTTCAAGATATCTACAGATGTAATCTTTGGGAAATATATAAAATCTGTCTCT 802  
 Db 2198 CAATTTCTTCAAGATATCTACAGATGTAATCTTTGGGAAATATATAAAATCTGTCTCT 2257  
 QY 803 GGAATACTAATGAAGAGAAAGTCCAGCAATTTGGAACTTTTATCATGAATTTCAATTAAT 862  
 Db 2258 GGAATACTAATGAAGAGAAAGTCCAGCAATTTGGAACTTTTATCATGAATTTCAATTAAT 2317  
 QY 863 TCTTGTGAGTCTCAAGAGCTTAACCAAGTAGACCAAGTACCACTTCTCAGCCCTACT 922  
 Db 2318 TCTTGTGAGTCTCAAGAGCTTAACCAAGTAGACCAAGTACCACTTCTCAGCCCTACT 2377  
 QY 923 TCAGAAACCAACAGAGCTTTTCAATGATCTCTTTTAAATAATATCAAAACCGTGTGAA 982  
 Db 2378 TCAGAAACCAACAGAGCTTTTCAATGATCTCTTTTAAATAATATCAAAACCGTGTGAA 2437  
 QY 983 AATACAGAGAGAAAGCAAAATCAGAAATTTATTTATGGAGGTGATGATCTGTAGAGGAGTC 1042

RESULT 8  
 AK000362 3145 bp mRNA linear PRI 22-FEB-2000  
 LOCUS Homo sapiens cDNA FLJ20355 fis, clone HEP15804, highly similar to  
 DEFINITION AF121863 Homo sapiens sorting nexin 14.  
 ACCESSION AK000362  
 VERSION AK000362.1 GI:7020397  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,  
 Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,  
 Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,  
 Nakamura,Y., Isogai,T. and Sugano,S.  
 NEBO human cDNA sequencing project  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3145)  
 AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
 Shibahara,T., Tanaka,T. and Nakamura,Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,  
 University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1,

Db 2438 AATACAGAGAGAAAGCAAAATCAGAAATTTATTTATGGAGGTGATGACTCTAGAGGAGTC 2497  
 QY 1043 TATGATTACCTGATGATGTAGGACGGGTAGTTTCCAGGTTCTCAGCTGGCTTCTATCAT 1102  
 Db 2498 TATGATTACCTGATGATGTAGGACGGGTAGTTTCCAGATTCTCAGCTGGCTTCTATCAT 2557  
 QY 1103 CTCTTAATGGAACTCGAATCTCTTTAAAAACACCCCTCGGAAATGTATCTGATTACTAT 1162  
 Db 2558 CTCTTAATGGAACTCGAATCTCTTTAAAAACACCCCTCGGAAATGTATCTGATTACTAT 2617  
 QY 1163 CTTCAAGTAACTAGAACAGCTATTTCCAGGAGCACCGTTTGGTCTCAGCTCATAAACATT 1222  
 Db 2618 CTTCAAGTAACTAGAACAGCTATTTCCAGGAGCACCGTTTGGTCTCAGCTCATAAACATT 2677  
 QY 1223 CTTCAAGATCTATATTTCTGTGAAACACCTGACCTCGCTCTCTCCCAAGATAAGCAAAA 1282  
 Db 2678 CTTCAAGATCTATATTTCTGTGAAACACCTGACCTCGCTCTCTCCCAAGATAAGCAAAA 2737  
 QY 1283 GAGCAAAAACAGACTTTTGAAGAAATGATGAATTTACATTTCCAGATCTGTTAGTCAAGTGT 1342  
 Db 2738 GAGCAAAAACAGACTTTTGAAGAAATGATGAATTTACATTTCCAGATCTGTTAGTCAAGTGT 2797  
 QY 1343 ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTTCTGTTGATGGCTTTACAGCAACCA 1402  
 Db 2798 ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTTCTGTTGATGGCTTTACAGCAACCA 2857  
 QY 1403 GTACTCAACAGCAGCTGACTATGTTTATTTGGAATGTGTGATCAGGAACTGTTTCCA 1462  
 Db 2858 GTACTCAACAGCAGCTGACTATGTTTATTTGGAATGTGTGATCAGGAACTGTTTCCA 2917  
 QY 1463 GAGCTCAATGAAGTACAAAAGGAAGTTACCTCTGTGACATCTTTGGATGTAAACACTTGA 1522  
 Db 2918 GAGCTCAATGAAGTACAAAAGGAAGTTACCTCTGTGACATCTTTGGATGTAAACACTTGA 2977  
 QY 1523 TTTGGTATAGATAAACCAATTTGCAATTTCTGCTGTGCGAGGGTGTAGAAATTTACTTTT 1582  
 Db 2978 TTTGGTATAGATAAACCAATTTGCAATTTCTGCTGTGCGAGGGTGTAGAAATTTACTTTT 3037  
 QY 1583 TTGGGTATATCTTATATATATTTATGTACATCCCTGTCTGAAATTTTAGTATTTTGT 1642  
 Db 3038 TTGGGTATATCTTATATATATTTATGTACATCCCTGTCTGAAATTTTAGTATTTTGT 3096  
 QY 1643 TTTTAAATAGACTAACAACAACTTAATGATTAATA 1678  
 Db 3097 TTTTAAATAGACTAACAACAACTTAATGATTAATA 3131



Minato-ku, Tokyo 108-8639, Japan (E-mail: cdna@ims.u-tokyo.ac.jp,  
Tel: 81-3-5449-5286, Fax: 81-3-5449-5416)  
NEBO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction; 5'- & 3'-end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

## FEATURES

## source

Location/Qualifiers  
1. 3145  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HEP15804"  
/cell\_line="HepG2"  
/cell\_type="hepatoma"  
/clone\_lib="HBP"  
/note="Cloning vector pME18SPL3"

## misc\_feature

1. 3145  
/note="highly similar to AF121863 Homo sapiens sorting  
nexin 14"

BASE COUNT 985 a 570 c 659 g 931 t

## ORIGIN

Query Match 75.5%; Score 1504; DB 9; Length 3145;  
Best Local Similarity 99.5%; Pred. No. 7e-256;  
Matches 1529; Conservative 0; Mismatches 5; Indels 2; Gaps 2;  
QY 143 TTCAGGACACACAGAAAGGGAGATCATTTGGAAATCAGCAGAAATGAGTACCAAAAT 202  
DB 1598 TTTCGGAACACACAGAAAGGGAGATCATTTGGAAATCAGCAGAAATGAGTACCAAAAT 1657  
QY 203 AAAGAGATTTCAAAGTACCACAAATGGAGGAGCTATGTTGCCATAATTAGGTGTAGCT 262  
DB 1658 AAAGAGATTTCAAGAGTACCACAAATGGAGGAGCTATGTTGCCATAATTAGGTGTAGCT 1717  
QY 263 GAAGTGAAGATGATTTTATTAAGAGAGTATTTGTAATGGAAGATGATTTCCAGTG 322  
DB 1718 GAAGTGAAGATGATTTTATTAAGAGAGTATTTGTAATGGAAGATGATTTCCAGTG 1777  
QY 323 GAGGCTGTGACACACCTTAATCTCCCGAACTTCTGTCATGCGAAATAGCAATCCA 382  
DB 1778 GAGGCTGTGACACACCTTAATCTCCCGAACTTCTGTCATGCGAAATAGCAATCCA 1837  
QY 383 TATGTAGACTTTTGTAGGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTCCT 442  
DB 1838 TATGTAGACTTTTGTAGGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTCCT 1897  
QY 443 GTGTTTCTGATTTGATTTGAAGAAATGATAGAGAGCAGTTGGACACGAGCTGGAACAT 502  
DB 1898 GTGTTTCTGATTTGATTTGAAGAAATGATAGAGAGCAGTTGGACACGAGCTGGAACAT 1957  
QY 503 TGGTCTGTCTATAGAGATATCTTGAATCTATGTATCTTGAATCAAAACTTAACAGATTT 562  
DB 1958 TGGTCTGTCTATAGAGATATCTTGAATCTATGTATCTTGAATCAAAACTTAACAGATTT 2017  
QY 563 CATGGTGATTTCTGTGATCCAGCTTCTCTTAAGAGATCATTTGCCCCCAAAATTTAT 622  
DB 2018 CATGGTGATTTCTGTGATCCAGCTTCTCTTAAGAGATCATTTGCCCCCAAAATTTAT 2077  
QY 623 GAATTTCTTAAGTCAAAGAGGAGAGTTCCAGAAATATCTACAGAACTTCTGCAGCAT 682  
DB 2078 GAATTTCTTAAGTCAAAGAGGAGAGTTCCAGAAATATCTACAGAACTTCTGCAGCAT 2137  
QY 683 CCAGAACTGAGTAATAGTCAATCTCTGGCAGACTTTCTTCCCTTAATGTTGGGGAACA 742  
DB 2138 CCAGAACTGAGTAATAGTCAATCTCTGGCAGACTTTCTTCCCTTAATGTTGGGGAACA 2197  
QY 743 CAATTTCTTGATAAGATCTACCATCTAATCTTGGGAAATATATAAATCTGTTCCT 802  
DB 2198 CAATTTCTTGATAAGATCTACCATCTAATCTTGGGAAATATATAAATCTGTTCCT 2257

QY 803 GGAACCTAATGAAGAGAAAGGTGAGCATTTGGAACCTTTTATCATGAATTTCAATTAAT 862  
DB 2258 GGAACCTAATGAAGAGAAAGGTGAGCATTTGGAACCTTTTATCATGAATTTCAATTAAT 2317  
QY 863 TCTTGTAGTCTCCAAAGCCTAAACCAAGTAGAGACAGAACTGAGCAATCTCAGCCCTACT 922  
DB 2318 TCTTGTAGTCTCCAAAGCCTAAACCAAGTAGAGACAGAACTGAGCAATCTCAGCCCTACT 2377  
QY 923 TCAGAAACACACAGAAAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGAA 982  
DB 2378 TCAGAAACACACAGAAAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGAA 2437  
QY 983 AATACAGAGAGAAAGCAAAATCAGAAATTTATTTATGAGGTGATGATCTAGAGAGGATC 1042  
DB 2438 AATACAGAGAGAGAAAGCAAAATCAGAAATTTATTTATGAGGTGATGATCTAGAGAGGATC 2497  
QY 1043 TATGATTAACCTGATGATAGTAGGACGGGTAGTTTCCAGGTTCCTGACTGCTCATCAT 1102  
DB 2498 TATGATTAACCTGATGATAGTAGGACGGGTAGTTTCCAGATTCCTGACTGCTCATCAT 2557  
QY 1103 CTCCTTAATGGGAACCTGGAATCTCTTTAAACACCCCTGGAATGTATATCTGATCTAT 1162  
DB 2558 CTCCTTAATGGGAACCTGGAATCTCTTTAAACACCCCTGGAATGTATATCTGATCTAT 2617  
QY 1163 CTTTCAGTGTAACTAGAAACAGCTATTTTCAGGACACCGTTTGTCTCACTCATACACTT 1222  
DB 2618 CTTTCAGTGTAACTAGAAACAGCTATTTTCAGGACACCGTTTGTCTCACTCATACACTT 2677  
QY 1223 CTCAGAGATGCTATATTTCTGTGAAACACAGCTGAACTCTCTCCAGAGTAAGCAAAA 1282  
DB 2678 CTCAGAGATGCTATATTTCTGTGAAACACAGCTGAACTCTCTCCAGAGTAAGCAAAA 2737  
QY 1283 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTTACATTCAGATCTGTTAGTCAAGTGT 1342  
DB 2738 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTTACATTCAGATCTGTTAGTCAAGTGT 2797  
QY 1343 ATTGTGTAGGAAACCAAGTATGAAGCATCAGCTTCTGTTGTGATGGCTTACAGCAACA 1402  
DB 2798 ATTGTGTAGGAAACCAAGTATGAAGCATCAGCTTCTGTTGTGATGGCTTACAGCAACA 2857  
QY 1403 GTACTCAACAGCAGCTGACTATTTTATTTGACATTTGATGATACAGGAACTGTTTCCA 1462  
DB 2858 GTACTCAACAGCAGCTGACTATTTTATTTGACATTTGATGATACAGGAACTGTTTCCA 2917  
QY 1463 GAGCTCAATAGGTACAAAGAGAACTTACCTCTGTGACATCTTGGATGTAAACACTTGG 1522  
DB 2918 GAGCTCAATAGGTACAAAGAGAACTTACCTCTGTGACATCTTGGATGTAAACACTTGG 2977  
QY 1523 TTTGGTATAGATAACCCATTTGAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTT 1582  
DB 2978 TTTGGTATAGATAACCCATTTGAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTT 3037  
QY 1583 TTTGGTATAGATAACCCATTTGAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTT 1642  
DB 3038 TTTGGTATAGATAACCCATTTGAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTT 3096  
QY 1643 TTTTAAATAAGACTAACCAAACTTTAAATGATTAAA 1678  
DB 3097 TTTTAAATAAGACTAACCAAC-ACCTAATGATTAAA 3131

## RESULT 9

## LOCUS

AY044865

2661 bp mRNA linear PRI 09-SEP-2001

Homo sapiens sorting nexin 14 (SNX14) mRNA, complete cds.

DEFINITION

AY044865

AY044865.1 GI:15529063

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 2661)



```
RESULT 10
BC043328
LOCUS
DEFINITION
MUS musculus RIKEN cDNA C330035N22 gene, mRNA (CDNA clone MGC:49424
IMAGE:5400157), complete cds.
ACCESSION
BC043328
VERSION
BC043328.1
KEYWORDS
GI:27694048
SOURCE
MGC.
ORGANISM
Mus musculus (house mouse)
MUS musculus
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1782)
Straussberg, R.D., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, F.S., Wagner, L., Shenmen, C.M., Schuler, G.D.,
Altschul, S.F., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Heide, F.,
Diatchenko, L., Marusina, K., Farmer, A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Schmitt, T.E., Brownstein, M.J., Uedin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullany, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettner, M., Madan, A., Rodriguez, S.,
Sanchez, A., Whitting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Bouffard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalilus, D.E.,
Schnerch, A., Schein, J.E., Jones, S.J., and Marra, M.A.
human and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
2238257
12477932
2 (bases 1 to 1782)
Straussberg, R.
Direct Submission
Submitted (09-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saedi, Jacqueline
Schein, Duane Smalilus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Teal, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 86 Row: P Column: 11
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
Location/Qualifiers
1..1782
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="MGC:49424 IMAGE:5400157"
Source
```

```
/tissue_type="Eye, retina, mouse strain C57BL/6"
/clone_lib="NIH_MGC_94"
/lab_host="DH10B"
/notes="Vector: PCMV-SPORT6"
1..1782
/gene="C330035N22Rik"
/notes="synonym: B830022K16"
/db_xref="LOCUSID:244962"
/db_xref="MGI:2443156"
228..1616
/codon_start=1
/product="C330035N22Rik protein"
/protein_id="AAH43328.1"
/db_xref="GI:27694049"
/db_xref="LOCUSID:244962"
/translation="MRICLIFRSTQKRGESFGISGRISKIKGVFKSTMEGAVLPNYG
VAEGDDTIERGIVVMEEDSVEAVSTENPRNLAANKISIPYVDFPDPSSRKEKK
BIPKVCIDVERNDRAVRHPEHPSVYRVLEFVLESKLTSEFGTQFDALPKRI
IGPKVYELKSKREFOVLOKLVHPELSNOLLADLSPNGGGTQFDLPLDVLN
GLIKSVGKLMKEKQHLPEFMSFINSCEPKPKRPELTLTILSPSENNKLPND
LFPNNANRAENTKQNYFMVENVGVYDLYMYGVFVQVFDLWHLHLMGRIL
FKNTLEMTDYLOCKLQELFQEHRLVSLITLLRDAIFCENTEPRSLQDKQKQTF
BEMNYIPDLIVKICIGETKVESIRLLPDLQQLPVNLKQLYVLLDIVIQELFPELNL
VQKEATSMTSMW"
BASE COUNT 567 a 339 c 380 g 496 t
ORIGIN
Query Match 59.9%; Score 1193.2; DB 10; Length 1782;
Best Local Similarity 86.5%; Pred. No. 6e-201;
Matches 1374; Conservative 0; Mismatches 183; Indels 31; Gaps 4;
79 TGTGTTTCAATAAAGTAGACTCTGTTGTGATAATTTGCTATGACTTGTGATACATTTTGT 138
182 TGGCATTGCTATAGTAGTCGGTGGTAAATTTGCTATGACTGATGATGCGTATTGTTT 241
139 AATATTCAGGACACACAGAAAGGGGGAATCATTTGGAATCAGACAGATAGTAGCAA 198
242 AATATTTAAGACACACAAAAAAGAGAGAAATCGTTTGAATCAGACAGATAGTAGCAA 301
199 AATTAAGAGGATATTAAGAAAGTACCAATGAGGAGGAGCTATGTTGCTTAATATGGTGT 258
302 AATTAAGGCGTATTCAGAGTACGACATGAGGAGGAGCTGCTGCTTAATATCGGGGT 361
259 AGCTGAAGGTGAAGATGATTTTATGAAGAGGATTTGTTGTAATGGAAGATGATTCCT 318
362 GCGTGAAGCGGAAGATGACTTTATTAAGAGAGGATTTGTTGTAATGGAAGATGACTCTCC 421
319 AGTGAAGGCTGTGAGCACACCTTAATCTCCCGAACCTTGTGCTGATGAAAATTAGCAT 378
422 AGTAGAAGCTGTGAGCACACCTTAACACTCTCGAACCTTGTGCTGGAATTTAGCAT 481
379 TCCATATGTAGACTTTTGTGAGGATCCCTCTCGAAAGGAGGAGAAAAAGAAAGAAAT 438
482 TCGTATGATGACTTTTGTGAAGATCCCTCTCGAAAGGAGGAGAGAGAGAGAT 541
439 TCGTGTGTTTGTATTGATGTTTGAAGAAATGATAGAAGAGAGAGTGGACAGACCTGA 498
542 TCGTGTGTTTGTATCGATGTTTGAAGAAACACGAGAGAGAGTGGACAGACCTGA 601
499 ACATTGGTCTGTATAGAAGATATCTTGAATTTCTATGATTTGAATCAAACTAACAGA 558
602 GCATTGGTCTGTATAGAAGATATCTTGAATTTCTATGATTTGAATCAAACTAACAGA 661
559 ATTTCAGTGTGATTTCTGATGCCAGCTTCTTCTTAAGAGGATCATTTGGCCCCCAAAA 618
662 ATTTCAGGACATTTCTGATGCTCAGCTTCCATCCAAAGAGATCATTTGGCCCCCAAAA 721
619 TTATGAATTTTAAAGTCAAGAGGAGAGATTTCCAAAGATATCTACGAATTTCTGCA 678
722 CTATGATTTTGAAGTCGAGAGAGAGAGATTTCCAGAGATATCTCGAGAACTTGTGCA 781
679 GCATCCAGAACTGAGTAATAGTCAACTTCTGCGACAGCTTTCTTCCCTTAATGGTGGGA 738
```

782 GCACCGGAGCTGAGTAACAGCCAGCTTCTGGCTGACTTTCTCCCCCAACGGTGGGA 841  
 739 AACACATTTCTGTAAAGTACTACAGATGTAAATCTTGGGAAAATTTATAAATCTGT 798  
 842 GACACAGTTTCTAGATAAGTACTTCCAGATGTAAATCTTGGGAAAATTTATAAATCTGT 901  
 799 TCTCGGAAAATTAATGAAGAGAAAGGTGAGATTTGGAACTTTTATCATGAATTTCA 858  
 902 TCTCGGAAAATTAATGAAGAGAAAGGTGAGATTTGGAACTTTTATCATGAATTTCA 961  
 859 TAATTTCTGTGAGTCTCCAAAGCTTAAACCAAGTAGACCAAGTGAACCATTTCTCAGGCC 918  
 962 TAATTTCTGTGAGTCTCCAAAGCTTAAACCAAGTAGACCAAGTGAACCATTTCTCAGGCC 1021  
 919 TACTTCAGAAAACACAGAGAGCTTTCAATGATCTGTTTAAATAATGCAAAACGGTGC 978  
 1022 CACTTCGAAAACACAGAGAGCTTTTCAACGATCTGTTTAAATAATGCAAAACGGTGC 1081  
 979 TGAATAATACAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTAGAAG 1038  
 1082 TGAGAACACAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTGACGG 1141  
 1039 AGTCTATGATTAACCTGATGTATGAGAGCGGTGATTTTCCAGTTCCTGACTGCTTCA 1098  
 1142 AGTCTATGATTAACCTGATGTATGAGAGCGGTGATTTTCCAGTTCCTGACTGCTTCA 1201  
 1099 TCATCTCTTAATGGGAATCTGAAATCTCTTTAAACACCTTGGAAAATGATATGATTA 1158  
 1202 TCATCTCTTAATGGGAATCTGAAATCTCTTTAAACACCTTGGAAAATGATATGATTA 1261  
 1159 CTATCTTCAGTGTAAATGAGACAGCTATTTCCAGAGACCGTTGGTCTCACTCATAC 1218  
 1262 CTACCTCCAGTGTGAGAGCGGTGATTTTCCAGAGACCGGTGGTCTCCTCATAC 1321  
 1219 ACTTCTCAGAGATGCTATATTTCTGTAACACACTGAACTCGTCTCTCCAAAGATAAGCA 1278  
 1322 CCTTCTCAGAGATGCTATATTTTGTGAATACTGAACTCGTCTCTCCAAAGATAAGCA 1381  
 1279 AAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAA 1338  
 1382 AAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAA 1441  
 1339 GTGTATTTGGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCA 1398  
 1442 GTGTATCGTGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCA 1501  
 1399 ACCAGTACTCAACAGCAGCTGACTTATGTTTATTTGACATTTGATGATACAGAACTGTT 1458  
 1502 GCCAGTCTCAACAGCAGCTGACTTATGTTTATTTGACATTTGATGATACAGAACTGTT 1561  
 1459 TCCAGAGCTCAATAAGGTACAAAAGGAAGTTACTCTGTGACATCTTGGATGTAAACACT 1518  
 1562 TCCAGAGCTCAATAAGGTACAAAAGGAAGTTACTCTGTGACATCTTGGATGTAAACACT 1620  
 1519 TGGATTTGATGAATAAACCAATTTGAAATTTCTGTGCGAGGTGTTAGAAATTTAC 1578  
 1621 TTTGATGATGAATAAACCAATTTGAAATTTCTGTGCGAGGTGTTAGAAATTTAC 1658  
 1579 TTTTCTGATATTTCTTATATATATATATATGATGATGATGATGATGATGATGATGAT 1636  
 1659 TTTTCTGATATTTTATATATATATATATGATGATGATGATGATGATGATGATGAT 1712  
 1637 TTTTGTGTTTAAAGAACTTAAACAAA 1664  
 1713 TTTTGTGTTTAAAGAACTTAAACAAA 1740

RESULT 11

LOCUS HSA420561 968 bp mRNA linear PRI 23-NOV-2001  
 DEFINITION Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 1521079.  
 ACCESSION AJ420561  
 VERSION AJ420561.1 GI:17066425

KEYWORDS Homo sapiens (human)  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1  
 AUTHORS Auffray, C., Anorge, W., Ballabio, A., Estivill, X., Gibson, K.,  
 Lehrach, H., Poustka, A. and Lundeberg, J.  
 TITLE The European IMAGE consortium for integrated Molecular analysis of  
 human gene transcripts  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 968)  
 AUTHORS Persson, A.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-OCT-2001) Persson A., Center for Molecular  
 Biotechnology, KTH, SCFAB, Institute of Biotechnology, Roslagssvagen  
 30B, 106 91 Stockholm, SWEDEN  
 COMMENT This clone is available royalty-free through IMAGE Consortium  
 Distributors. IMPORTANT: This sequence represents the full insert  
 of this IMAGE cDNA clone. No attempt has been made to verify  
 whether this corresponds to the full-length of the original mRNA  
 from which it was derived.  
 FEATURES  
 source  
 1. 968  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="EUROIMAGE 1521079"  
 /clone\_lib="NCI CGAP Lu5"  
 BASE COUNT 311 a 147 c 186 g - 324 t  
 ORIGIN  
 Query Match 46.5%; Score 925.4; DB 9; Length 968;  
 Best Local Similarity 99.8%; Pred. No. 1.3e-153;  
 Matches 937; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
 QY 1019 GAGTGTGATGCTGTAGAGGAGTCTATGATTAACCTGATGTATGTAGGACGGTAGTTTC 1078  
 Db 18 GGGGTGATGATGTAGAGGAGTCTATGATTAACCTGATGTATGTAGGACGGTAGTTTC 77  
 QY 1079 CAGGTTCCTGATCGCTTCAATCATCTCTTAATGGAACCTGAACTCTTTTAAACACAC 1138  
 Db 78 CAGGTTCCTGATCGCTTCAATCATCTCTTAATGGAACCTGAACTCTTTTAAACACAC 137  
 QY 1139 CTGAAATGTATATCTGATTAATCTTTCAGTGTAACTAGAACAGCTATTTTCAGGAGCAC 1198  
 Db 138 CTGAAATGTATATCTGATTAATCTTTCAGTGTAACTAGAACAGCTATTTTCAGGAGCAC 197  
 QY 1199 CGTTTGGTCTCACTAATAACATTTCTCAGAGATGCTATATTTCTGTGAAAACACTGAACTT 1258  
 Db 198 CGTTTGGTCTCACTAATAACATTTCTCAGAGATGCTATATTTCTGTGAAAACACTGAACTT 257  
 QY 1259 CGCTCTCTCCAAAGATAAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAATTAAC 1318  
 Db 258 CGCTCTCTCCAAAGATAAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAATTAAC 317  
 QY 1319 ATTCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAGATGAAAGCATCAGACTT 1378  
 Db 318 ATTCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAGATGAAAGCATCAGACTT 377  
 QY 1379 CTGTTTGTATGCTTACAGCAACCAAGTACTCAACAGCAGCTGATTAATGTTTATTTGAC 1438  
 Db 378 CTGTTTGTATGCTTACAGCAACCAAGTACTCAACAGCAGCTGATTAATGTTTATTTGAC 437  
 QY 1439 ATTGTGTATACAGGAACCTGTTTCCAGAGCTCAATAGGTACAAAGGAGTACCTCTCTGTG 1498  
 Db 438 ATTGTGTATACAGGAACCTGTTTCCAGAGCTCAATAGGTACAAAGGAGTACCTCTCTGTG 497  
 QY 1499 ACATCTTGGATGAACACCTTGGATTTGGTATATAAATTAACATTAACCAATTTGAAATTTTCGCTGTG 1558  
 Db 498 ACATCTTGGATGAACACCTTGGATTTGGTATATAAATTAACATTAACCAATTTGAAATTTTCGCTGTG 557







PC C12N15/09, C12N15/09, C01A15/41, C01A16/16, C12N1/13, C12N1/19, C12N1/21, C12N1/21,

PC C12N15/09, C12N15/09, C01A15/41, C01A16/16, C12N1/13, C12N1/19, C12N1/21, C12N1/21,



PI JORDAN  
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC  
C12N5/10.  
PC C12P21/02, C12P21/08, C12Q1/68//G06F17/30, C12N15/00, C12N5/00, PC  
G06F15/40

CC  
FH Key Location/Qualifiers  
FT CDS 53..358.

FEATURES  
source 1..358  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"

BASE COUNT 125 a -63 c -69 g 101 t

ORIGIN

Query Match 17.9%; Score 357; DB 6; Length 358;  
Best Local Similarity 100.0%; Pred. No. 3.7e-53;  
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 760 ACTACAGATGTAATCTTGGGAATTTATAAATCTGTCTCTGGAAACTAATGAAGA 819  
DB 1 ACTACAGATGTAATCTTGGGAATTTATAAATCTGTCTCTGGAAACTAATGAAGA 60  
QY 820 GAAAGGTCAGACTTGGAACTTTTATCATGAATTTTCAATTTCTTGAGTCTCCAA 879  
DB 61 GAAAGGTCAGACTTGGAACTTTTATCATGAATTTTCAATTTCTTGAGTCTCCAA 120  
QY 880 GCCTAAACCAAGTAGACCACTGACCACTTCTCAGCCCTACTTCAGAAAAACAAGAA 939  
DB 121 GCCTAAACCAAGTAGACCACTGACCACTTCTCAGCCCTACTTCAGAAAAACAAGAA 180  
QY 940 GCTTTTCAATGATCTGTTTAAATAATGCAACCGTGTGAAATACAGAGAAAGCA 999  
DB 181 GCTTTTCAATGATCTGTTTAAATAATGCAACCGTGTGAAATACAGAGAAAGCA 240  
QY 1000 AATATCAGATTTATTTATGAGGTGATGACTGTAGAGGAGTCTATGATTACCTGATGA 1059  
DB 241 AATATCAGATTTATTTATGAGGTGATGACTGTAGAGGAGTCTATGATTACCTGATGA 300  
QY 1060 TGTAGGACGGGTAGTTTCCAGGTTCTGACTGGCTTCATCTCTTAATGGGAAC 1116  
DB 301 TGTAGGACGGGTAGTTTCCAGGTTCTGACTGGCTTCATCTCTTAATGGGAAC 357

Search completed: January 31, 2004, 12:44:49  
Job time : 7419 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 10:23:38 : Search time 125 Seconds

(without alignments)  
7033.882 Million cell updates/sec

Title: US-09-744-313A-3

Perfect score: 1992

Sequence: 1 gctgaagactccaaagtt.....tttaataataaaaaaaaaa 1992

Scoring table: IDENTITY NUC

Gap: 1.0 Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.\*

- 1: /cgn2\_6/prodata/2/ina/5A COMB.seq.\*
- 2: /cgn2\_6/prodata/2/ina/5B COMB.seq.\*
- 3: /cgn2\_6/prodata/2/ina/6A COMB.seq.\*
- 4: /cgn2\_6/prodata/2/ina/6B COMB.seq.\*
- 5: /cgn2\_6/prodata/2/ina/PCTUS COMB.seq.\*
- 6: /cgn2\_6/prodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1832.8	76.9	1716	4	US-09-620-312D-922 Sequence 922, App
2	1392.8	59.9	1551	4	US-09-620-312D-923 Sequence 923, App
3	57.2	2.9	658	3	US-08-998-416-595 Sequence 595, App
4	53.8	2.7	7218	1	US-08-232-463-14 Sequence 14, App
5	53.2	2.7	8920	2	US-08-446-855A-1 Sequence 1, Appl
6	53.2	2.7	8920	3	US-09-150-741-1 Sequence 1, Appl
7	51.2	2.6	6124	4	US-08-213-419B-3 Sequence 3, Appl
8	48.2	2.4	2110	3	US-09-419-459-1 Sequence 1, Appl
9	47.4	2.4	615	3	US-08-998-416-186 Sequence 186, App
10	47.4	2.4	665	2	US-08-883-795A-36 Sequence 36, App
11	44.6	2.2	837	3	US-08-998-416-288 Sequence 288, App
12	44.4	2.2	6152	3	US-08-973-462-1 Sequence 1, Appl
13	44	2.2	1851	4	US-09-601-198-51 Sequence 51, Appl
14	43.8	2.2	483	4	US-09-220-132-18 Sequence 18, Appl
15	43.8	2.2	483	4	US-09-220-132-161 Sequence 161, App
16	43.8	2.2	662	3	US-08-998-416-185 Sequence 185, App
17	43.8	2.2	701	3	US-08-998-416-701 Sequence 701, App
18	43.8	2.2	724	3	US-08-998-416-683 Sequence 683, App
19	43.8	2.2	732	3	US-08-998-416-1036 Sequence 1036, App
20	43.8	2.2	767	3	US-08-998-416-472 Sequence 472, App
21	43.8	2.2	827	3	US-08-998-416-535 Sequence 535, App
22	43.8	2.2	828	3	US-08-998-416-538 Sequence 538, App
23	43.8	2.2	834	3	US-08-998-416-305 Sequence 305, App
24	43.8	2.2	10640	4	US-09-417-485B-5 Sequence 5, Appl
25	43.6	2.2	19124	2	US-08-487-826B-13 Sequence 13, Appl
26	43.4	2.2	636	3	US-08-998-416-1137 Sequence 1137, App
27	43.4	2.2	765	4	US-09-601-198-160 Sequence 160, App

C 28	43.4	2.2	1430	2	US-08-179-557-15	Sequence 15, Appl
29	43.2	2.2	1887	4	US-09-601-198-39	Sequence 39, Appl
30	43	2.2	1186	2	US-08-731-722-5	Sequence 5, Appl
C 31	42.8	2.1	148567	4	US-09-801-876B-3	Sequence 3, Appl
C 32	42.6	2.1	711	3	US-08-998-416-786	Sequence 786, App
C 33	42.4	2.1	1230025	4	US-09-198-452A-1	Sequence 1, Appl
C 34	42.4	2.1	665	3	US-08-998-416-937	Sequence 937, App
35	42.2	2.1	2030	2	US-08-705-937-7	Sequence 7, Appl
36	42.2	2.1	168575	4	US-09-426-290-1	Sequence 1, Appl
37	42	2.1	1497	4	US-09-620-312D-1021	Sequence 1021, App
38	42	2.1	6124	4	US-08-213-419B-3	Sequence 3, Appl
39	41.6	2.1	8920	2	US-08-446-855A-1	Sequence 1, Appl
40	41.6	2.1	8920	3	US-09-150-741-1	Sequence 1, Appl
41	41.6	2.1	19124	2	US-08-487-826B-13	Sequence 13, Appl
C 42	41.4	2.1	417	4	US-08-559-896B-3	Sequence 3, Appl
C 43	41.4	2.1	639	4	US-09-328-352-1706	Sequence 1706, App
C 44	41.4	2.1	1956	4	US-08-559-896B-1	Sequence 1, Appl
C 45	41.2	2.1	3060	4	US-09-996-243-504	Sequence 504, App

#### ALIGNMENTS

##### RESULT 1

US-09-620-312D-922  
; Sequence 922, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Mehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Wang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yuning  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes Version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (71)..(1543)

US-09-620-312D-922  
Query Match 76.9%; Score 1532.8; DB 4; Length 1716;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1534; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 143 TTCAGGACACACAGAAAACGGGAGGATCATTTGGATCAGCAGATAGTAGCAAAATT 202  
Db 173 TTTCGGAACACACAGAAAACGGGAGGATCATTTGGATCAGCAGATAGTAGCAAAATT 232

Qy	203	AAAGGAGTATTCAAAGTACCAATATGGAGGGAGCTAATGTTGCTTAATTAATGTTAGCT	262
Db	233	AAAGGAGTATTCAAAGTACCAATATGGAGGGAGCTAATGTTGCTTAATTAATGTTAGCT	292
Qy	263	GAAGGTGAAGATGATTTTATTTGAAGAGGTATTTGTTGTTAATGGAAGATGATTTCTCCAGTG	322
Db	293	GAAGGTGAAGATGATTTTATTTGAAGAGGTATTTGTTGTTAATGGAAGATGATTTCTCCAGTG	352
Qy	323	GAGGCTGTGAGCACACCTTAATCTCTCCCGGAAACCTTCTGCTGCATGGAAAAATTAGCATTTCCA	382
Db	353	GAGGCTGTGAGCACACCTTAATCTCTCCCGGAAACCTTCTGCTGCATGGAAAAATTAGCATTTCCA	412
Qy	383	TATGTAGACTTTTGTGAGATCCCTCTCTGAAAGGAGGAGAAAAAGAAAGAAATTCCT	442
Db	413	TATGTAGACTTTTGTGAGATCCCTCTCTGAAAGGAGGAGAAAAAGAAAGAAATTCCT	472
Qy	443	GTGTTTTGTATGATGTTGAAGAAATGATAGAGAGACAGTTGAGACACGAGCCCTGAGCAT	502
Db	473	GTGTTTTGTATGATGTTGAAGAAATGATAGAGAGACAGTTGAGACACGAGCCCTGAGCAT	532
Qy	503	TGGTCTCTCTATAGAAGATATCTTGAATTTCTATGTAATTTGAATCAAAACTTAAACAGAAATTT	562
Db	533	TGGTCTCTCTATAGAAGATATCTTGAATTTCTATGTAATTTGAATCAAAACTTAAACAGAAATTT	592
Qy	563	CATGGTGCATTTCTGTATGCCAGCTTCTTTCTAAGAGATCATTTGGCCCCAAAAATTTAT	622
Db	593	CATGGTGCATTTCTGTATGCCAGCTTCTTTCTAAGAGATCATTTGGCCCCAAAAATTTAT	652
Qy	623	GNATTTCTTAAGTCAAGAGGAGGAGAGTCCAGAAATATCTACAGAAATCTTCTGCGAGCAT	682
Db	653	GNATTTCTTAAGTCAAGAGGAGGAGAGTCCAGAAATATCTACAGAAATCTTCTGCGAGCAT	712
Qy	683	CCAGAACTGAGTAAATAGTCAACTTTCTGGCAGACTTTCTTTCCCTTAATGTTGGGGAAACA	742
Db	713	CCAGAACTGAGTAAATAGTCAACTTTCTGGCAGACTTTCTTTCCCTTAATGTTGGGGAAACA	772
Qy	743	CAATTTCTTGATTAAGATCTACAGATGTAAATCTTGGGAAATATATAAATCTGTTCTCT	802
Db	773	CAATTTCTTGATTAAGATCTACAGATGTAAATCTTGGGAAATATATAAATCTGTTCTCT	832
Qy	803	GGAAAACTAATGAAGAGAAAGGTCCAGATTTGGAACTTTTATCATGAATTTTCATTAAT	862
Db	833	GGAAAACTAATGAAGAGAAAGGTCCAGATTTGGAACTTTTATCATGAATTTTCATTAAT	892
Qy	863	TCTTGTGAGTCTCCAAAGCCTTAAACCAAGTAGACAGAACTGACCAATTTCTCAGCCCTACT	922
Db	893	TCTTGTGAGTCTCCAAAGCCTTAAACCAAGTAGACAGAACTGACCAATTTCTCAGCCCTACT	952
Qy	923	TCAGAAAAACAAGAGAGCTTTTCAATGATCTGTTTTAAAAAATTAATGAAACCGTGCTGAA	982
Db	953	TCAGAAAAACAAGAGAGCTTTTCAATGATCTGTTTTAAAAAATTAATGAAACCGTGCTGAA	1012
Qy	983	AATACAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTAGAAGGATC	1042
Db	1013	AATACAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTAGAAGGATC	1072
Qy	1043	TATGATTAACCTGATGATGTAGAGCGGTAGTTTTTCCAGGTTCCTGACTGGCTTCATCAT	1102
Db	1073	TATGATTAACCTGATGATGTAGAGCGGTAGTTTTTCCAGGTTCCTGACTGGCTTCATCAT	1132
Qy	1103	CTCTTAATGGGAATCTGAAATCTCTTTTAAAAACAACCTTGGAAATGTATATCTGATTAAT	1162
Db	1133	CTCTTAATGGGAATCTGAAATCTCTTTTAAAAACAACCTTGGAAATGTATATCTGATTAAT	1192
Qy	1163	CTTCAGGTGAACCTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTTCACTCATTAACATTT	1222
Db	1193	CTTCAGGTGAACCTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTTCACTCATTAACATTT	1252
Qy	1223	CTCAGAGATGCTATATTCTGTGAAAAACACTGAACTCGCTCTCTCCCAAGATAAGCAAAAA	1282
Db	1253	CTCAGAGATGCTATATTCTGTGAAAAACACTGAACTCGCTCTCTCCCAAGATAAGCAAAAA	1312
Qy	1283	GGAGCAAAAACAGACTTTTGTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGT	1342

## RESULT 2

```

US-09-620-312D-923
; Sequence 923, Application US/09620312D
; Patent No. 656962
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenchua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyang
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yungqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6369662el Nucleic Acids and
; Polypeptides
; FILE REFERENCE: 784CIP28
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/486,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 923
; LENGTH: 1551
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71)..(1378)
US-09-620-312D-923

```

Query Match 59.9%; Score 1192.8; DB 4; Length 1551;  
Best Local Similarity 89.1%; Pred. No. 6.8e-268;  
Matches 1369; Conservative 0; Mismatches 2; Indels 165;

143 TTCAAGAACACACAGAAAGGGGAGAAATCAATTTGGAAATCAGCAAGATAGTAGCAAAAT 202  
173 TTTCCGACACACAGAAAGGGGAGAAATCAATTTGGAAATCAGCAAGATAGTAGCAAAAT 232  
203 AAAGAGTATTCAAAAGTACCAAAATGAGGAGGATATGTTGCTTAATATGTTAGCT 262  
233 AAAGAGTATTCAAAAGTACCAAAATGAGGAGGATATGTTGCTTAATATGTTAGCT 292  
263 GAAGGTGAAGATGATTTTATTTGAAGAGGATATGTTGCTTAATATGTTAGCT 322  
293 GAAGGTGAAGATGATTTTATTTGAAGAGGATATGTTGCTTAATATGTTAGCT 352  
323 GAGGCTGAGGACACCACTTACTCCCGGAAACCTTGCTGATGAGGAAATAGCAATTC 382  
353 GAGGCTGAGGACACCACTTACTCCCGGAAACCTTGCTGATGAGGAAATAGCAATTC 412  
383 TATGTAGACTTTTGTAGGATCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGG 442  
413 TATGTAGACTTTTGTAGGATCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGG 472  
443 GTGTTTGTATGATGTTGAAAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 502  
473 GTGTTTGTATGATGTTGAAAGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGG 532  
503 TGTGCTGCTATAGAGATATCTGAAATCTTGAATCTGAAATCTGAAATCTGAAATCT 562  
533 TGTGCTGCTATAGAGATATCTGAAATCTTGAATCTGAAATCTGAAATCTGAAATCT 592  
563 CATGGTGCATTTCTGATGCTGAGCTTCTTCTTGAAGAGGATCAATGGCCGCCAAATTA 622  
593 CATGGTGCATTTCTGATGCTGAGCTTCTTCTTGAAGAGGATCAATGGCCGCCAAATTA 652  
623 GAATTTTAAAGTCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 682  
653 GAATTTTAAAGTCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 712  
683 CCAGAACTGAGTAAATAGTCAACTTCTGCGAGAGCTTTTCCCTTAATGTTGGGAGAAC 742  
713 CCAGAACTGAGTAAATAGTCAACTTCTGCGAGAGCTTTTCCCTTAATGTTGGGAGAAC 772  
743 CAATTTCTGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 802  
773 CAATTTCTGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 832  
803 GGAAACTAATGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 862  
833 GGAAACTAATGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 892  
863 TCTTGTGAGTCTCAAAGGCTTAAACCAAGTACCAAGTACCAAGTACCAAGTACCAAG 922  
893 TCTTGTGAGTCTCAAAGGCTTAAACCAAGTACCAAGTACCAAGTACCAAGTACCAAG 952  
923 TCAGAAACCAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 982  
953 TCAGAAACCAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1012  
983 AATACAG 1042  
1013 AATACAG 1072  
1043 TATGATTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1102  
1073 TATGATTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1092  
1103 CTCTTAATGGGAACCTGAAATCTCTTTAAAGAACACCCCTGGAATGATATGATTAAT 1162  
1093 ----- 1092  
1163 CTTCCAGTGTAACTAGBACAGCTATTTCCAGGAGACCGTTTGGTCTCTCATCAACACTT 1222  
1093 ----- 1092

1223 CTCAGAGATGCTATATTTCTGTGAAACACACGAACTCGCTCTCTCCAGGATAGCAAA 1282  
1093 -----AGATGCTATATTTCTGTGAAACACACGAACTCGCTCTCTCCAGGATAGCAAA 1147  
1283 CGAGCAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGT 1342  
1148 CGAGCAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGT 1207  
1343 ATTGGTGAAGAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGT 1402  
1208 ATTGGTGAAGAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGT 1267  
1403 GTACTCTAACAGAGAGCTGACTTTATGTTTATTTGACATTTGATGATGATGATGATGATGAT 1462  
1268 GTACTCTAACAGAGAGCTGACTTTATGTTTATTTGACATTTGATGATGATGATGATGATGAT 1327  
1463 GAGCTCAATAGAGTACAAAAGAGTACCTCTGTCGATGATGATGATGATGATGATGATGATGAT 1522  
1328 GAGCTCAATAGAGTACAAAAGAGTACCTCTGTCGATGATGATGATGATGATGATGATGATGAT 1387  
1523 TTTGGTATAGATTAACCCATTTGAAATTTCTGCTGTCGAGGAGGAGGAGGAGGAGGAGGAGG 1582  
1388 TTTGGTATAGATTAACCCATTTGAAATTTCTGCTGTCGAGGAGGAGGAGGAGGAGGAGGAGG 1447  
1583 TTTGGTATATTTCTTATATATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 1642  
1448 TTTGGTATATTTCTTATATATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 1507  
1643 TTTTAAATAAGACTAACACAACTTAATGATTAATA 1678  
1508 TTTTAAATAAGACTAACACAACTTAATGATTAATA 1543

## RESULT 3

US-08-998-416-595  
; Sequence 595, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippen, Peter  
; APPLICANT: Fohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Rebischung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPIL  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/998,416  
; FILING DATE: 24-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 0016/97  
; FILING DATE: 31-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587



```

RESULT 6
US-09-150-741-1/c
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02

```

RESULT 7  
 US-08-213-419B-3/C  
 ; Sequence 3, Application US/08213419B  
 ; Patent No. 6333406  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Inselburg, J. et al.  
 ; TITLE OF INVENTION: GENE ENCODING PROTEIN ANTIGENS OF PLASMODIUM FALCIPARUM  
 ; TITLE OF INVENTION: AND USES THEREFOR  
 ; FILE REFERENCE: J11-002CNCP  
 ; CURRENT APPLICATION NUMBER: US/08/213,419B  
 ; CURRENT FILING DATE: 1994-03-14  
 ; PRIOR APPLICATION NUMBER: US 07/870,506  
 ; PRIOR FILING DATE: 1992-04-17  
 ; NUMBER OF SEQ ID NOS: 20  
 ; SOFTWARE: PatentIn Ver. 2.0







QY 1752 GCATTTTAAAGAGAAATTCATAATGATGCTATGGCAACAGATAGACTGATAAC 1811  
DB 372 AATATTGTAATTTATTTTATTAATAATCTATTTTATAAATATTATGTTGATT 431  
QY 1812 TTGTAATGTAAGCTTTGAAATAATTAATGCTAGTATGGAGAAACAGGAATAGATCT 1871  
DB 432 ATATTATTAACTTTTATTAAGAAATTAATTATAAATTAATTTAACTTTTCTTAT 491  
QY 1872 GATTTCTTAGAGTAAATATTTTAGTAGATGCTTTTCTTTTATTTTATTTGTCAT 1931  
DB 492 TATTATTTTATATTTATTAATAATTAATTTATTTATTTATTTATTTATTTAATA 551  
QY 1932 AGTTAACTGCTATCATATAATAAAGCACTCCATATGATGTTTTTAATAATAAAAAAAA 1991  
DB 552 AATTAATTTATTAATTAATTTATTTATTTATTTATTTATTTATTTATTTAATA 611  
QY 1992 A 1992  
DB 612 A 612

RESULT 10  
US-08-883-795A-36/c  
; Sequence 36, Application US/08883795A  
; Patent No. 5985607  
; GENERAL INFORMATION:  
; APPLICANT: Delcuve, Genevieve  
; APPLICANT: Awang, Gregor  
; TITLE OF INVENTION: Recombinant DNA Molecules and Expression  
; TITLE OF INVENTION: Vectors for Tissue Plasminogen Activator  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BERSKIN & PARR  
; STREET: 40 King Street West  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5H 3Y2  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/883,795A  
; FILING DATE: 27-JUN-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Gravelle, Micheline  
; REGISTRATION NUMBER: 40,261  
; REFERENCE/DOCKET NUMBER: 7841-062  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 364-7311  
; TELEFAX: (416) 361-1398  
; INFORMATION FOR SEQ ID NO: 36:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 665 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cdna  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; IMMEDIATE SOURCE:  
; CLONE: Rh 32  
US-08-883-795A-36  
Query Match 2.48; Score 47.4; DB 2; Length 665;  
Best Local Similarity 47.28; Pred. No. 0.051;  
Matches 144; Conservative 0; Mismatches 161; Indels 0; Gaps 0;  
QY 1628 TTAGTATTTTGTGTTTTTAAAGACTAACCAACTTAATGATTAAGTGTAG 1687

DB 344 TTATAAACATTTTAAATTAATAATATGTAATTAATAACATTTTAAATATAAATATTAA 285  
QY 1688 TCCTAGTCTTTTCATTTGCTAGCTGATCCAAATTTTATTAGACATAAAGTCACTTGT 1747  
DB 284 TTATAAACATTTTAAATTAATAATATTTAAATTAATAATTTTAAATATAAATATTAA 225  
QY 1748 TATTGCCATTTTAAAGAGAAATTCATAATGATGCTTATGGCAACAGATAGACTGAT 1807  
DB 224 TTATAAATTTTAAATTAATAATATTTAAATTAATAATTTTAAATATAAATATTAA 165  
QY 1808 AACTTCGTATGTAAGCTTTGAAATAATTAATGCTAGTATGGAGAAACAGGAATAAG 1867  
DB 164 TTATAAATATTTTAAATTAATAATATTTAAATTAATAATTTTAAATATAAATATTAA 105  
QY 1868 ATCTGATTTTCTAGAGTTAATAATTTTAGTAGATGTTGTTTCTTTTATTTTGT 1927  
DB 104 TTATAAATATTTTAAATTAATAATATTTAAATTAATAATTTTAAATATAAATATTAA 45  
QY 1928 ACATA 1932  
DB 44 TTATA 40

RESULT 11  
US-08-998-416-288  
; Sequence 288, Application US/08998416  
; Patent No. 6239264  
; GENERAL INFORMATION:  
; APPLICANT: Philippsen, Peter  
; APPLICANT: Pohlmann, Rainer  
; APPLICANT: Steiner, Sabine  
; APPLICANT: Mohr, Christine  
; APPLICANT: Wendland, Jurgen  
; APPLICANT: Knechtle, Philipp  
; APPLICANT: Reibschung, Corinne  
; TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSYPII  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 1152  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 6239264artis Corporation  
; STREET: 3054 Cornwallis Road  
; CITY: Research Triangle Park  
; STATE: No. 6239264th Carolina  
; COUNTRY: USA  
; ZIP: 27709  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/998,416  
; FILING DATE: 24-DEC-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: CH 0016/97  
; FILING DATE: 31-DEC-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Meigs, J. Timothy  
; REGISTRATION NUMBER: 38,241  
; REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 919-541-8587  
; TELEFAX: 919-541-8689  
; INFORMATION FOR SEQ ID NO: 288:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 837 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; ORIGINAL SOURCE:  
; ORGANISM: PAG1241RP

US-08-998-416-288

Query Match 2.2%; Score 44.6; DB 3; Length 837;  
Best Local Similarity 47.3%; Pred. No. 0.24; Mismatches 169; Conservative 0; Gaps 1;  
Matches 169; Conservative 0; Mismatches 184; Indels 4; Gaps 1;  
QY 1632 TTAATTTTGGTTTAAATAGAGCTAAACACAACTTAATGATTAAGAGTGAATGAGTCTC 1691  
DB TTAATAGATATATATTTCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 311  
QY 1692 AFAGCTTTCTAGTCTGAGTGCATTAATTTTATTTATTTAGAACATAAGTCTGTTTATT 1751  
DB TTAATTAATTAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTT 371  
QY 1752 GCCATTTTAAAGAGAAATTCATTAATGATGTTATGCGCAACAGATAGACTGATTAAC 1811  
DB AAATATGTAATTAATTTTATTAATTAATTTATTTTATTAATTAATTTATTTATTTGATTT 431  
QY 1812 TFCGATTTAGTATGAGTTTGAATAATTAATGCTAGTATGAGAAACAGGAATAAGATCT 1871  
DB ATATTTTAACTTTT-----ATTAAGATTTATTTAAATTAATTTAACTTTAAATTT 487  
QY 1872 GATTTTCTAGAGTTAAATATATTTTAGTAGATGTTTCTTTTCTTTTATTTTGTACAT 1931  
DB CTATTTATTAATTTTATTTATTTAAATTAATTTATTTATTTATTTATTTATTTATTT 547  
QY 1932 AGTTAACTGATCTATTAATTAAGATCCATATAGATTTTAAATTAATTAATTAATTA 1988  
DB TAAATTAATTAATTTTAAATTAATTTATTTATTTATTTATTTATTTATTTATTTATTA 604

## RESULT 12

US-08-973-462-1/c  
; Sequence 1, Application US/08973462B  
; Patent No. 6191270  
; GENERAL INFORMATION:  
; APPLICANT: DRUILHE, PIERRE  
; APPLICANT: DAUBERTS, PIERRE  
; TITLE OF INVENTION: MAJALAL PRE-BRYTHROCYTIC STAGE POLYPEPTIDE MOLECULES  
; FILE REFERENCE: 0660-0125-0 PCT  
; CURRENT APPLICATION NUMBER: US/08/973,462B  
; EARLIER FILING DATE: 1998-02-06  
; EARLIER APPLICATION NUMBER: PCT/FR96/00894  
; EARLIER FILING DATE: 1996-06-12  
; EARLIER APPLICATION NUMBER: FR 95/07007  
; EARLIER FILING DATE: 1995-06-13  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 6152  
; TYPE: DNA  
; ORGANISM: P. falciparum  
US-08-973-462-1

Query Match 2.2%; Score 44.4; DB 3; Length 6152;  
Best Local Similarity 48.7%; Pred. No. 0.55;  
Matches 149; Conservative 0; Mismatches 156; Indels 1; Gaps 1;  
QY 1623 AAATTTTAGTATTTTGTGTTTAAATAGACTAAACACAACTTAATGATTAAGTGA 1682  
DB AT 5912  
QY 1683 TTGAGTCTCATAGTCTTTTCAATTTGCTAGCTGATCCAAATTTTATTTAGAACATAAGTCA 1742  
DB TATGTAATGAATAATATACAGTTTCTTACAGAAATCTATATATTAATTAATTAATTAATTA 5852  
QY 1743 CTGTTATTGCCATTTTAAAGAGAAATTCATATGATGTTATGCGCAACAGATAGA 1802  
DB CTCTGTAAACAACTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5792  
QY 1803 CTGATAACTTCGTATTGTATGATCTTTGAAATTAATTTATGCTAGTATGAGAAACAGGA 1862  
DB AATAAATATTAATTAATTAATTAATTTTAAATTTTAAATTTTAAATTTTAAATTTTAAAT 5732

QY 1863 ATA-AGATCTGATTTTCTTAGAGTTAAATATATATTTTAGTAGATTTGGTTTCTTTTATTA 1921  
DB ATA 5672  
QY 1922 TTTTGT 1927  
DB CATTTT 5666

## RESULT 13

US-09-601-198-51  
; Sequence 51, Application US/09601198  
; Patent No. 6531583  
; GENERAL INFORMATION:  
; APPLICANT: Cassell, Gail H.  
; APPLICANT: Chen, Ellison Y.  
; APPLICANT: Glass, Jennifer S.  
; APPLICANT: Glass, John I.  
; APPLICANT: Heiner, Cheryl R.  
; APPLICANT: Lefkowitz, Elliot  
; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA  
; FILE REFERENCE: UAB-13452/22  
; CURRENT APPLICATION NUMBER: US/09/601,198  
; CURRENT FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/073,189  
; PRIOR FILING DATE: 1998-01-30  
; NUMBER OF SEQ ID NOS: 181  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 51  
; LENGTH: 1851  
; TYPE: DNA  
; ORGANISM: Ureaplasma urealyticum  
US-09-601-198-51

Query Match 2.2%; Score 44; DB 4; Length 1851;  
Best Local Similarity 45.7%; Pred. No. 0.44;  
Matches 190; Conservative 0; Mismatches 225; Indels 1; Gaps 1;

QY 1573 ATTACTTTTTGGGTATATCTTATATATATATATATATATATATATATATATATATAT 1632  
DB ATGTTCAATTTTGTGTAATATTTTGTCTTCAATTAATTTCTTAGCTTTGGG 213  
QY 1633 TATTTTGTGTTTTTAAAGAACTAAACAACTTAATTAATTAATTAATTAATTAATTAAT 1692  
DB AATATAAGCTTTTATCAATCATAGTAAGAACTCTAACACTTAAGATTTTCAAGTTCTTT 273  
QY 1693 TAGTCTTTCATTTGCTAG-CTGTGATCCAAATTTTATTTAGAACATAAGTCTGTTATT 1751  
DB GCTAAATCTCTTCTTGTGATTTATTTAAATCTAAATAGCGATTAATTTCTTGCGTTT 333  
QY 1752 GCCATTTTAAAGAGAGAAATTCATAATGATGTTATGCAACACAGATAAGCTGATAAC 1811  
DB TTGACGGAATCAGAGCAATATGTTAAATCAATTTTTCAGATTTATTTACTGTTAATAC 393  
QY 1812 TTGCTATTGCTAGCTTTGAAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 1871  
DB TTGCTTTGAGCAAGTTTAAATTAATTTAGTATTTTAAATAAATCTTTTTCATACT 453  
QY 1872 GATTTTCTTAGAGTTAAATATTTTATTTAGTAGATGTTTCTTTTCTTTTATTTTGTACAT 1931  
DB TTTTAAGACTCTTCTGAATGTTTGAATTCGATTTGATTTTATTTATTTGATCGTTTTCATT 513  
QY 1932 AGTTAACTGATCTATTAATTAAGCATCTTATATGAGTTTATTAATTAATTAATTAAT 1987  
DB ATTTAATCAATTAATTAATCTTTTAAATGTTTAAATGAGTACTTAAATAATAATAA 569

## RESULT 14

US-09-220-132-18/c  
; Sequence 18, Application US/09220132  
; Patent No. 6506607

GENERAL INFORMATION:  
APPLICANT: Shyjan, Andrew W.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT  
OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 07334-074001  
CURRENT APPLICATION NUMBER: US/09/220,132  
CURRENT FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: US 60/079,303  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: US 60/068,821  
PRIOR FILING DATE: 1997-12-24  
NUMBER OF SEQ ID NOS: 191  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 18  
LENGTH: 483  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(483)  
OTHER INFORMATION: n = A, T, C or G  
US-09-220-132-18

Query Match 2.2%; Score 43.8; DB 4; Length 483;  
Best Local Similarity 54.4%; Pred. No. 0.31;  
Matches 87; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 1822 ATAGCTTTGAAATAAATTATGCTAGTAGTGGAGAAACAGGAATAAGATCTGATTTCTTA 1881  
DB 392 ATGGGTTTAAATTAAAGAGCATCCGGTTTGGTATGGGNATGATCCAGGATATGTTG 333  
QY 1882 GAGTTAATATATTTAGTAGATGTTGTTCCCTTTTATTTTGTACATAGTAACTGT 1941  
DB 332 TGACTGATACATATAGTTACTTGTGCTTTTTTTTTTTTTTTTGGATCTTTGCAAGGC 273  
QY 1942 GTATCTATAATAAAGCATCCCTATATGAGTTTATAATAAT 1981  
DB 272 AAACACAGTAACGAGTTTATATAATAATTAAATT 233

RESULT 15  
US-09-220-132-161/C  
Sequence 161, Application US/09220132  
Patent No. 6506607  
GENERAL INFORMATION:  
APPLICANT: Shyjan, Andrew W.  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT  
OF PROSTATE CANCER THERAPIES AND THE DIAGNOSIS OF PROSTATE CANCER  
FILE REFERENCE: 07334-074001  
CURRENT APPLICATION NUMBER: US/09/220,132  
CURRENT FILING DATE: 1998-12-23  
PRIOR APPLICATION NUMBER: US 60/079,303  
PRIOR FILING DATE: 1998-03-25  
PRIOR APPLICATION NUMBER: US 60/068,821  
PRIOR FILING DATE: 1997-12-24  
NUMBER OF SEQ ID NOS: 191  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 161  
LENGTH: 483  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (1)...(483)  
OTHER INFORMATION: n = A, T, C or G  
US-09-220-132-161

Query Match 2.2%; Score 43.8; DB 4; Length 483;  
Best Local Similarity 54.4%; Pred. No. 0.31;  
Matches 87; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
QY 1822 ATAGCTTTGAAATAAATTATGCTAGTAGTGGAGAAACAGGAATAAGATCTGATTTCTTA 1881

DB 392 ATGGGTTTAAATAAAGAGCATCCGGTTTGGTATGGGNATGATCCAGGATATGTTG 333  
QY 1882 GAGTTAATATATTTAGTAGATGTTTCCCTTTTATTTTGTACATAGTAACTGT 1941  
DB 332 TGACTGATACATATAGTTACTTGTGCTTTTTTTTTTTTGGATCTTTGCAAGGC 273  
QY 1942 GTATCTATAATAAAGCATCCCTATATGAGTTTATAATAAT 1981  
DB 272 AAACACAGTAACGAGTTTATATAATAATTAAATT 233

Search completed: January 31, 2004, 13:58:47  
Job time : 130 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 09:03:22 ; Search time 547 Seconds  
(without alignments)  
9830.495 Million cell updates/sec

Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: 1-gtataaaactccaaagt.....tttaataataaaaaaaaa 1992

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2552756-seq, 134219017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

- Database : N\_Geneseq\_19Jun03.\*
- 1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
  - 2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
  - 3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
  - 4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
  - 5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
  - 6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
  - 7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
  - 8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
  - 9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
  - 10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
  - 11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
  - 12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
  - 13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
  - 14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
  - 15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
  - 16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
  - 17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
  - 18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
  - 19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
  - 20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
  - 21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
  - 22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
  - 23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
  - 24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*
  - 25: /SIDSL1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1992	100.0	1992	22 AAC86397	SNEXN DNA #1. Hom
C 2	1840.2	92.4	2176	22 AAI60815	Human polynucleoti
C 3	1840.2	92.4	2176	22 AAI60816	Human polynucleoti
4	1532.8	76.9	1716	22 AAI59029	Human polynucleoti
5	1504	75.5	3145	24 ABQ95348	CDNA encoding a pr
6	1192.8	59.9	1551	22 AAI59030	Human polynucleoti
7	655.4	32.9	779	20 AAZ17379	Human gene express
C 8	427.2	21.4	451	24 ABL63556	Breast cancer rela

9	424.4	21.3	725	24	ABS77486	Frog embryonic gen
10	364.2	18.3	402	25	ABX50611	Bovine EST associa
11	357	17.9	358	21	AAC02755	Human secreted pro
12	267.2	13.4	305	25	ABX50426	Bovine EST associa
13	86	4.3	311	22	AAH69552	Human cervical can
14	86	4.3	313	22	AAH70887	Human cervical can
15	84	4.2	299	22	AAH72445	Human cervical can
16	84	4.2	299	22	AAH73007	Human cervical can
C 17	78.6	3.9	580	24	ABN62287	Human cancer relat
18	68.4	3.4	9095	24	ABQ67061	Human angiogenesis
C 19	67.2	3.4	8056	25	ABZ10246	Haematopoietic cel
20	66.2	3.3	7461	24	ABL33785	Human immune syste
21	65.6	3.3	6161	24	ABL32623	Human immune syste
C 22	65.6	3.3	8056	25	ABZ10100	Haematopoietic cel
23	63.4	3.2	7306	24	ABN80291	Human chemically m
24	63.4	3.2	18133	24	ABK40017	Human chemically p
25	63.4	3.2	18133	24	ABL32940	Human immune syste
26	62.6	3.1	4590	7	AAH60472	Sequence encoding
27	62.2	3.1	73334	24	ABL92319	Chemically treated
28	62.2	3.1	73334	24	ABL34125	Human immune syste
29	62	3.1	14316	24	ABL70605	Chemically treated
30	62	3.1	14316	24	AAS61444	Human gene regulat
31	62	3.1	14316	24	ABK31518	Signal transductio
32	61.6	3.1	15767	24	ABL33207	Human immune syste
33	61.6	3.1	15767	24	ABL34553	Human metastasis a
34	61.6	3.1	17183	24	ABL32487	Human immune syste
35	61.4	3.1	5997	24	ABL33624	Human immune syste
36	60.8	3.1	8056	25	ABZ10246	Haematopoietic cel
37	60.6	3.0	5413	22	AAS46694	Tumour suppressor
38	60.2	3.0	61020	22	AAS46788	Tumour suppressor
39	60	3.0	60	60	ABN39970	Human spliced tran
40	60	3.0	6090	25	ABZ10150	Haematopoietic cel
41	59.8	3.0	6203	22	AAS45476	Chemically pretrea
42	59.8	3.0	6203	24	ABK28400	DNA transcription
C 43	59.4	3.0	612	22	AAH71471	Human cervical can
44	59.4	3.0	9145	24	ABL32888	Human immune syste
45	59.4	3.0	17389	24	ABL33415	Human immune syste

ALIGNMENTS

RESULT 1  
AAC86397

ID AAC86397 standard; DNA; 1992 BP.

XX AC AAC86397;

XX DT 21-MAR-2001 (first entry)

XX DE SNEXN DNA #1.

XX	XX	XX	XX	XX	XX	SNEXN; human; sorting nexin; inflammation; asthma; allergy; AIDS; neurological disorder; gastrointestinal; smooth muscle cell; cancer; gene therapy; ds.
XX	OS	XX	XX	XX	XX	Homo sapiens.
XX	PN	XX	XX	XX	XX	WO200073334-A2.
XX	PD	XX	XX	XX	XX	07-DEC-2000.
XX	PP	XX	XX	XX	XX	26-MAY-2000; 2000WO-US14831.
XX	PR	XX	XX	XX	XX	27-MAY-1999; 98US-0136740.
XX	PA	XX	XX	XX	XX	16-JUN-1999; 99US-0139566.
XX	PI	XX	XX	XX	XX	(INCV-) INCYTE GENOMICS INC.
XX	PI	XX	XX	XX	XX	Yue H, Tang YT, Azimzai Y;
XX	DR	XX	XX	XX	XX	WPI; 2001-041141/05.
XX	XX	XX	XX	XX	XX	

PT Novel Sorting Nexin polypeptides and polynucleotides useful for  
 PT diagnosing and treating disorders associated with their expression e.g.  
 PT autoimmune disorders, smooth muscle cell disorders and cell  
 PT proliferative disorders

PS Claim 5; Page 82-83; 84pp; English.

XX The present invention relates to human sorting nexin (SNEXN).  
 CC Compositions containing SNEXN or agonists of SNEXN are useful for  
 CC treating a disease or condition associated with decreased expression of  
 CC functional SNEXN and compositions containing antagonists of SNEXN are  
 CC useful for treating a disease or condition associated with  
 CC overexpression of functional SNEXN. These can be immune  
 CC disorders for example inflammation, asthma, allergy, and AIDS,  
 CC neurological disorders, gastrointestinal disorders, smooth muscle cell  
 CC disorders, cancers and others. The SNEXN may also be used for somatic  
 CC or germline gene therapy.

SQ Sequence 1992 BP; 671 A; 308 C; 372 G; 641 T; 0 other;

Query Match 100.0%; Score 1992; DB 22; Length 1992;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GTATGAAACTCCAAAGTTATGTAGCTATACCTTTAGTTTATCATTTTCAAACTGTT	60
DB	1	GTATGAAACTCCAAAGTTATGTAGCTATACCTTTAGTTTATCATTTTCAAACTGTT	60
QY	61	TTTCTTTTTTTTAAATGTTTTTCAATAAACTAGACTGTGTGATAAATTTGCTATGTA	120
DB	61	TTTCTTTTTTTTAAATGTTTTTCAATAAACTAGACTGTGTGATAAATTTGCTATGTA	120
QY	121	CTTGATACATTTTGTGTTTAAATTCAGGAACAACAGAAAAGGGGAGAAATCATTTGGAAT	180
DB	121	CTTGATACATTTTGTGTTTAAATTCAGGAACAACAGAAAAGGGGAGAAATCATTTGGAAT	180
QY	181	CAGCAGATAGTAGCAAAATTAAGAGATTTCAAAAGTACCACAAATGGAGGAGCTAT	240
DB	181	CAGCAGATAGTAGCAAAATTAAGAGATTTCAAAAGTACCACAAATGGAGGAGCTAT	240
QY	241	GTTCCTTAATTTGGTGTAGCTGAGGTGAAGTGAAGTGAATTTTGAAGAGGTATTGTTGT	300
DB	241	GTTCCTTAATTTGGTGTAGCTGAGGTGAAGTGAAGTGAATTTTGAAGAGGTATTGTTGT	300
QY	301	AATGGAAGATGATTTCTCCAGTGGAGGCTGTGAGCACACCTAATCTCCCGAAACCTTGC	360
DB	301	AATGGAAGATGATTTCTCCAGTGGAGGCTGTGAGCACACCTAATCTCCCGAAACCTTGC	360
QY	361	TGCATGGAATAATAGCAATCCATATGATAGACTTTTGGAGATCCCTCTCTGAAAGGAA	420
DB	361	TGCATGGAATAATAGCAATCCATATGATAGACTTTTGGAGATCCCTCTCTGAAAGGAA	420
QY	421	GGAGAAAAGAAAGAAATTCCTGTGTTTGTATGATGTTGAAGAAATGATAGAGAGC	480
DB	421	GGAGAAAAGAAAGAAATTCCTGTGTTTGTATGATGTTGAAGAAATGATAGAGAGC	480
QY	481	AGTTGACACGAGCCTGAAATGCTGTCTATAGAGATATCTTGAATTTCTATGTAAT	540
DB	481	AGTTGACACGAGCCTGAAATGCTGTCTATAGAGATATCTTGAATTTCTATGTAAT	540
QY	541	TGAATCAAACTAACAGAAATTTCAATGTCATTTCTGATGCCAGCTTCTTCTTAAGAG	600
DB	541	TGAATCAAACTAACAGAAATTTCAATGTCATTTCTGATGCCAGCTTCTTCTTAAGAG	600
QY	601	GATCATTGGCCCCCAAAATTAAGAAATTTAAAGTCAAGAGAGGAGAGTCCCAAGATA	660
DB	601	GATCATTGGCCCCCAAAATTAAGAAATTTAAAGTCAAGAGAGGAGAGTCCCAAGATA	660
QY	661	TCTACAGAACTTCTGAGCATCCAGAACTGAGTAATAGTCAACTTCTGGCAGACTTCT	720
DB	661	TCTACAGAACTTCTGAGCATCCAGAACTGAGTAATAGTCAACTTCTGGCAGACTTCT	720
QY	721	TTCCCTTAATGGTGGGAAACACAATTTCTTGTATAGATACCTACAGATGTAATCTTGG	780

DB	721	TTCCCTTAATGGTGGGAAACACAATTTCTTGTATAGATACCTACAGATGTAATCTTGG	780
QY	781	GAAATTAATAAATCTGTTCTGGAAAACCTAATGAAGAGAAAGTCAAGCTTTGGAACC	840
DB	781	GAAATTAATAAATCTGTTCTGGAAAACCTAATGAAGAGAAAGTCAAGCTTTGGAACC	840
QY	841	TTTTATCATGAATTTCAATTAATTTCTGTGAGTCTCCAAAGCCTAAACCAAGTAGACCAGA	900
DB	841	TTTTATCATGAATTTCAATTAATTTCTGTGAGTCTCCAAAGCCTAAACCAAGTAGACCAGA	900
QY	901	ACTGACCATCTCTCAGCCCTTACTTTCAGAAAACCAACAGAGCTTTTCAATGATCTGTTAA	960
DB	901	ACTGACCATCTCTCAGCCCTTACTTTCAGAAAACCAACAGAGCTTTTCAATGATCTGTTAA	960
QY	961	AAATAATCAAAACCGTGTCTGAAAAATACAGAGAGAAAGCAAAATCAGAAATTTATGGA	1020
DB	961	AAATAATCAAAACCGTGTCTGAAAAATACAGAGAGAAAGCAAAATCAGAAATTTATGGA	1020
QY	1021	GGTGATGCTGTAGAGGAGTCTATGATTTACCTGATGATGTAGGCGGTAGTTTCCA	1080
DB	1021	GGTGATGCTGTAGAGGAGTCTATGATTTACCTGATGATGTAGGCGGTAGTTTCCA	1080
QY	1081	GGTTCCTGACTGGCTTCATCATCTCTTAATGGAACTCGAATCCCTCTTTAAAAACACCCCT	1140
DB	1081	GGTTCCTGACTGGCTTCATCATCTCTTAATGGAACTCGAATCCCTCTTTAAAAACACCCCT	1140
QY	1141	GGAAATGATATGATTTACTTCTCAGTGTAAACCTAGAAACAGCTATTTTCAGAGCAGCG	1200
DB	1141	GGAAATGATATGATTTACTTCTCAGTGTAAACCTAGAAACAGCTATTTTCAGAGCAGCG	1200
QY	1201	TTTGGTCTCACTCATTAACACITTCAGAGATGCTATATTTCTGTGAAAACACTGAACTCG	1260
DB	1201	TTTGGTCTCACTCATTAACACITTCAGAGATGCTATATTTCTGTGAAAACACTGAACTCG	1260
QY	1261	CTCTCTCCAGATPAAGCAAAAGGAGCAAAAACAGACTTTTGAAGAAATCATGAATTACAT	1320
DB	1261	CTCTCTCCAGATPAAGCAAAAGGAGCAAAAACAGACTTTTGAAGAAATCATGAATTACAT	1320
QY	1321	TCAGATCTGTTTGTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCT	1380
DB	1321	TCAGATCTGTTTGTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCT	1380
QY	1381	GTTCGATGCTTACAGAAACAGATCTCAACAGAGCTGACTATGTTTATTTGACAT	1440
DB	1381	GTTCGATGCTTACAGAAACAGATCTCAACAGAGCTGACTATGTTTATTTGACAT	1440
QY	1441	TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAAGAAAGTTACCTCTGTGAC	1500
DB	1441	TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAAGAAAGTTACCTCTGTGAC	1500
QY	1501	ATCTTGGATGTAAACACTTGGATTTGGTATAGAAATTAACCCATTTGAAATTTCTGCTGTGC	1560
DB	1501	ATCTTGGATGTAAACACTTGGATTTGGTATAGAAATTAACCCATTTGAAATTTCTGCTGTGC	1560
QY	1561	AGGTGGTAGAAATTTACTTTTTTGGGTATATTTCTTATATATATATGATCATCGCTGTC	1620
DB	1561	AGGTGGTAGAAATTTACTTTTTTGGGTATATTTCTTATATATATATGATCATCGCTGTC	1620
QY	1621	TGAAATTTTGTATTTTGTGTTTTTAAATAAGACTTAACACAACTTAATGATTTAAAGT	1680
DB	1621	TGAAATTTTGTATTTTGTGTTTTTAAATAAGACTTAACACAACTTAATGATTTAAAGT	1680
QY	1681	GATTGAGTCTCATAGTCTTTCATTTCTGTAGCTGTGATCCAAATTTTATTTAGAACATAAGT	1740
DB	1681	GATTGAGTCTCATAGTCTTTCATTTCTGTAGCTGTGATCCAAATTTTATTTAGAACATAAGT	1740
QY	1741	CACCTGTTATTTGCCATTTTAAAGAGAAATTCATAATGATGTTTATGGCAACAGATAA	1800
DB	1741	CACCTGTTATTTGCCATTTTAAAGAGAAATTCATAATGATGTTTATGGCAACAGATAA	1800
QY	1801	GACTGATAAATCTGTTATGTTATAGCTTTGAAAATTAATTTATGCTAGTATGGAGAACAG	1860

Db 1801 GACTGATAAACTCGTATTGTATAGCTTTGAAATAAATTATGCTAGTATGGAGAAACAG 1860  
Qy 1861 GAAATAAGCTGATTTCTTAGAGTAAATATATTTTAGTAGATTGGTTTTCCTTTT 1920  
Db 1861 GAAATAAGCTGATTTCTTAGAGTAAATATATTTTAGTAGATTGGTTTTCCTTTT 1920  
Qy 1921 ATTTTGTACATAGTAACTGTTGATCTATATAAATAAAGCATCCTATATGAGTTTATAA 1980  
Db 1921 ATTTTGTACATAGTAACTGTTGATCTATATAAATAAAGCATCCTATATGAGTTTATAA 1980  
Qy 1981 TAAAAA 1992  
Db 1981 TAAAAA 1992  
RESULT 2  
AAI60815/c  
ID AAI60815 standard; cDNA; 2176 BP.  
AC AAI60815;  
XX  
DT 22-OCT-2001 (first entry)  
XX Human polynucleotide SEQ ID NO 4804.  
DE Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic;  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX  
XX 25-APR-2000; 2000US-0552317.  
XX  
XX 09-JUL-2000; 2000US-0598042.  
XX  
XX 19-JUL-2000; 2000US-0620312.  
XX  
XX 03-AUG-2000; 2000US-0653450.  
XX  
XX 14-SEP-2000; 2000US-0662191.  
XX  
XX 19-OCT-2000; 2000US-0693036.  
XX  
XX 29-NOV-2000; 2000US-0727344.  
XX  
XX (HYSE-) HYSEQ INC.  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
XX Zhao Q, Zhou P, Goodrich R, Drmanac RT;  
XX  
XX WPI; 2001-442253/47.  
XX  
XX P-PSDB; AAM41659.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders.  
XX such as central nervous system injuries -  
XX  
XX Claim 1; SEQ ID NO 4804; 10078pp; English.  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
XX immunosuppressant and cytostatic activity. The polynucleotides are useful  
XX in gene therapy. A composition containing a polypeptide or polynucleotide  
XX of the invention may be used to treat diseases of the peripheral nervous  
XX system, such as peripheral nervous injuries, peripheral neuropathy and  
XX localised neuropathies and central nervous system diseases, such as  
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
XX utilisation of the activities such as: Immune system suppression,

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 2176 BP; 684 A; 427 C; 353 G; 712 T; 0 other;  
Query Match 92.4%; Score 1840.2; DB 22; Length 2176;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1842; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 143 TTCAGGAACACACAGAAAAGGGAGATCATTTTGGAAATCAGCAGATAGTACAAAATT 202  
Db 2022 TTTCGGAACACACAGAAAAGGGAGATCATTTTGGAAATCAGCAGATAGTACAAAATT 1963  
Qy 203 AAAGGAGTATTCAAAAGTACCACAAATGAGGAGCTATGTTGCCCTAAATATGGTGTAGCT 262  
Db 1962 AAAGGAGTATTCAAAAGTACCACAAATGAGGAGCTATGTTGCCCTAAATATGGTGTAGCT 1903  
Qy 263 GAAGGTGAAGATGATTTTATTAAGAGAGGTATTTGTAAGGAGATGATTCCTCAGTG 322  
Db 1902 GAAGGTGAAGATGATTTTATTAAGAGAGGTATTTGTAAGGAGATGATTCCTCAGTG 1843  
Qy 323 GAGGCTGTGAGCACACACCTAATACTCCCGAAACCTTGTGCTGATGAAAATTAGCAATCCA 382  
Db 1842 GAGGCTGTGAGCACACACCTAATACTCCCGAAACCTTGTGCTGATGAAAATTAGCAATCCA 1783  
Qy 383 TATGTAGACTTTTGTGAGGATCCCTCTCTGAAAGGAGGAGAGAAAAGAAAGAAATTCCT 442  
Db 1782 TATGTAGACTTTTGTGAGGATCCCTCTCTGAAAGGAGGAGAGAAAAGAAAGAAATTCCT 1723  
Qy 443 GTGTTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACAGCGCTGAAACAT 502  
Db 1722 GTGTTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACAGCGCTGAAACAT 1663  
Qy 503 TGGTCTGTCTATAGAGATATCTTGAATCTATGTACTTGAATCAAAACTAACAGAAATTT 562  
Db 1662 TGGTCTGTCTATAGAGATATCTTGAATCTATGTACTTGAATCAAAACTAACAGAAATTT 1603  
Qy 563 CATGTGCTATTTCTGATGCCAGCTTCTTCTAAGAGGATCATTTGGCCCCCAAAAATTTAT 622  
Db 1602 CATGTGCTATTTCTGATGCCAGCTTCTTCTAAGAGGATCATTTGGCCCCCAAAAATTTAT 1543  
Qy 623 GAAATTCCTTAAAGTCAAGAGGAGGAGAGTTCCAAGAAATATCTACAGAAATCTTTCGAGCAT 682  
Db 1542 GAAATTCCTTAAAGTCAAGAGGAGGAGAGTTCCAAGAAATATCTACAGAAATCTTTCGAGCAT 1483  
Qy 683 CCAGAACTGAGTAAATAGTCAACTTCTGCGAGACTTTCTTCCCTTAATGGTGGGGAACA 742  
Db 1482 CCAGAACTGAGTAAATAGTCAACTTCTGCGAGACTTTCTTCCCTTAATGGTGGGGAACA 1423  
Qy 743 CAATTTCTTGATAGATATCTACAGATGTAATCTTGGGAAATTTATAAATCTGTTCTCT 802  
Db 1422 CAATTTCTTGATAGATATCTACAGATGTAATCTTGGGAAATTTATAAATCTGTTCTCT 1363  
Qy 803 GGAATACTAATGAAAGAGAAAGGTGAGCAATTTGAAACCTTTTATCATCAATTTCAATTAAT 862  
Db 1362 GGAATACTAATGAAAGAGAAAGGTGAGCAATTTGAAACCTTTTATCATCAATTTCAATTAAT 1303  
Qy 863 TCTTGTGAGTCTCCAAAGCCTTAAACCAAGTAGACCCAGAACTGACCATTTCTCAGCCCTACT 922  
Db 1302 TCTTGTGAGTCTCCAAAGCCTTAAACCAAGTAGACCCAGAACTGACCATTTCTCAGCCCTACT 1243  
Qy 923 TCAGAAAACACACAGAGCTTTTCAATCATCTGTTTAAATAATAGCAAAACGGTGTGAA 982  
Db 1242 TCAGAAAACACACAGAGCTTTTCAATCATCTGTTTAAATAATAGCAAAACGGTGTGAA 1183  
Qy 983 AATACAGAGAGAAACCAAAATCAGAAATTTTATGAGAGTGAATGACTGTAGAGAGGTC 1042  
Db 1182 AATACAGAGAGAAACCAAAATCAGAAATTTTATGAGAGTGAATGACTGTAGAGAGGTC 1123







QY	1403	GTACTCAACAGCAGCTGACTTATGTATTTATTTGGACATTTGTGATACAGGAACGTGTTCCA	1466
DB	762	GTACTCAACAGCAGCTGACTTATGTATTTATTTGGACATTTGTGATACAGGAACGTGTTCCA	703
QY	1463	GAGCTCAATAAGGTACAAAGGAAGTTACCTCTGTGACATCTTTGGATGTAACACATTGGA	1522
DB	702	GAGCTCAATAAGGTACAAAGGAAGTTACCTCTGTGACATCTTTGGATGTAACACATTGGA	643
QY	1523	TTTGGTATAGAATAACCCCAATTCGAAATTTCTGCTGTGCGAGGGTGGTAGAAATTTTACTTTT	1582
DB	642	TTTGGTATAGAATAACCCCAATTCGAAATTTCTGCTGTGCGAGGGTGGTAGAAATTTTACTTTT	583
QY	1583	TTGGGTATATTCCTTATATATATATATGATACATCGCTGCTCGAAATTTTATAGTATATTTTGT	1642
DB	582	TTGGGTATATTCCTTATATATATATATGATACATCGCTGCTCGAAATTTTATAGTATATTTTGT	523
QY	1643	TTTTTAATAAGACTAACACAAACTTAATGATTTAAAAGTGAATTGAGTCTCATAGTCTTTTCA	1702
DB	522	TTTTTAATAAGACTAACACAAACTTAATGATTTAAAAGTGAATTGAGTCTCATAGTCTTTTCA	463
QY	1703	TTTGCTAGCTGTGATCCAAATTTTATTAGAACATAAGTCACTTGTATTGCCCATTTTTTAA	1762
DB	462	TTTGCTAGCTGTGATCCAAATTTTATTAGAACATAAGTCACTTGTATTGCCCATTTTTTAA	403
QY	1763	AAGAGAAAATTCATAATGATGTTTATGCGCAACACAGATAAGATGATAAACTTCGTATTGTA	1822
DB	402	AAGAGAAAATTCATAATGATGTTTATGCGCAACACAGATAAGATGATAAACTTCGTATTGTA	343
QY	1823	TAGCTTTGAAAATAATATATGCTTAGTATGAGAGAAACAGGAATAAGATCTGTATTTCTTAG	1882
DB	342	TAGCTTTGAAAATAATATATGCTTAGTATGAGAGAAACAGGAATAAGATCTGTATTTCTTAG	283
QY	1883	AGTTAATATATTTTATAGTAGATTGGTTTTCTTTTTTTTTTATTTTGTACATAGTTAACTGTG	1942
DB	282	AGTTAATATATTTTATAGTAGATTGGTTTTCTTTTTTTTTTATTTTGTACATAGTTAACTGTG	223
QY	1943	TATCTATAAATAAGCATCCTATATGATGATTTTTTAAATAATAAAAAA	1987
DB	222	TATCTATAAATAAGCATCCTATATGATGATTTTTTAAATAATAAAAAA	178
RESULT 4			
AAI59029			
ID	AAI59029 standard; cDNA; 1716 BP.		
XX	AAI59029;		
XX	22-OCT-2001 (first entry)		
XX	Human polynucleotide SEQ ID NO 1232.		
KW	Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
XX	leukaemia; ss.		
OS	Homo sapiens.		
XX	WO200153312-A1.		
XX	26-JUL-2001.		
XX	26-DEC-2000; 2000WO-US34263.		
XX	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		

RESULT 4  
AAI59029

ID AAI59029 standard; cDNA; 1716 bp.

AAI59029:

DT : 22-OCT-2001 (first entry)

Human polynucleotide SEQ ID NO 1232.

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; KW  
peripheral nervous system; neuropathy; central nervous system; CNS; KW  
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; KW  
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic; KW  
chemokinetic; thrombolytic; drug screening; arthritis; inflammation; KW  
leukaemia; ss. KW

Homocapiens.

XX  
PN  
W0200153312-A1

26-ПТ.-2001-

26-DEC-2000: 2000WO-US34263-XX PF

21--JAN-2000: 2000US-0488725.

PK 23-APK-2000; 2000US-0332317.  
 PR 09-III-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-062031Z.  
PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR	19-OCT-2000; 2000US-0691036.	Db	653	GAATTTCTTAAAGTCAAGAGGGAAGAGTTCACAGAAATATCTACAGAAATCTTCGACGAT	712
PR	29-NOV-2000; 2000US-0727344.	Qy	683	CCAGAACTGAGTAATAGTCAACTCTGCGACACTCTTCTCCCTTAATGGTGGGGAACA	742
XX	(HYSE-) HYSEQ INC.	Db	713	CCAGAACTGAGTAATAGTCAACTCTGCGACACTCTTCTCCCTTAATGGTGGGGAACA	772
XX	Tang Y, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	Qy	743	CAATTTCTTGATAGATATACTACAGATGTAATCTTGGGAAATATATAAAATCTGTTCT	802
XX	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	Db	773	CAATTTCTTGATAGATATACTACAGATGTAATCTTGGGAAATATATAAAATCTGTTCT	832
XX	Zhao Q, Zhou P, Goodrich R, Drmanac RT;	Qy	803	GGAAACTTAATGAAGAGAAAGGTGAGATTTGGAACTTTTATCATGAAATTCATTAAT	862
DR	WPI; 2001-442253/47.	Db	833	GGAAACTTAATGAAGAGAAAGGTGAGATTTGGAACTTTTATCATGAAATTCATTAAT	892
XX	P-PSDB; AAM39873.	Qy	863	TCCTTGAGTCTCCAAAGCCTAAACCAAGTAGAGACAGAACTGACCATCTCAGCCCTACT	922
PT	Novel nucleic acids and polypeptides, useful for treating disorders	Db	893	TCCTTGAGTCTCCAAAGCCTAAACCAAGTAGAGACAGAACTGACCATCTCAGCCCTACT	952
XX	such as central nervous system injuries -	Qy	923	TCAGAAACCAACCAAGAGCTTTTCAATGATCTGTTTAAAAAATAATGCAAAACCGTCTGAA	982
PS	Claim 1; SEQ ID NO 1232; 10078pp; English.	Db	953	TCAGAAACCAACCAAGAGCTTTTCAATGATCTGTTTAAAAAATAATGCAAAACCGTCTGAA	1012
XX	The invention relates to human nucleic acids (AAI57798-AAI61369) and	Qy	983	AATACAGAGAGAAACCAAAATCAGAAATTTATTTATGGAGGTGATGACTGTAGAAGAGTC	1042
CC	the encoded polypeptides (AAM38642-AM42213) with nootropic,	Db	1013	AATACAGAGAGAAACCAAAATCAGAAATTTATTTATGGAGGTGATGACTGTAGAAGAGTC	1072
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	Qy	1043	TATGATTACTGATGATAGAGAGGTAGTTTCCAGGTTTCCAGGTTTCCAGGTTTCCAGGTTT	1102
CC	in gene therapy. A composition containing a polypeptide or polynucleotide	Db	1073	TATGATTACTGATGATAGAGAGGTAGTTTCCAGGTTTCCAGGTTTCCAGGTTTCCAGGTTT	1132
CC	of the invention may be used to treat diseases of the peripheral nervous	Qy	1103	CTCTTAATGGGAACCTCGAATCTCTTTAAAAACACCTTGGAAATCTATATGATTACTAT	1162
CC	system, such as peripheral nervous injuries, peripheral neuropathy and	Db	1133	CTCTTAATGGGAACCTCGAATCTCTTTAAAAACACCTTGGAAATCTATATGATTACTAT	1192
CC	localised neuropathies and central nervous system diseases, such as	Qy	1163	CTTTCAGTGAAGTAACTAGAACAGCTATTTACAGAGACCGTTTGGTCTCAGTCAATACACTT	1222
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	Db	1193	CTTTCAGTGAAGTAACTAGAACAGCTATTTACAGAGACCGTTTGGTCTCAGTCAATACACTT	1252
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	Qy	1223	CTCAGAGATGCTATATTTCTGTGAAACACTGAACCTCGCTCTCCAGATTAAGCAAAA	1282
CC	utilisation of the activities such as: immune system suppression,	Db	1253	CTCAGAGATGCTATATTTCTGTGAAACACTGAACCTCGCTCTCCAGATTAAGCAAAA	1312
CC	Activ/inhibin activity, chemotactic/chemokinetic activity, haemostatic	Qy	1283	GGAGCAAAACAGACTTTTTCGAAGAAATGATGAATTTACATTCAGATCTGTTAGTCAAGTGT	1342
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	Db	1313	GGAGCAAAACAGACTTTTTCGAAGAAATGATGAATTTACATTCAGATCTGTTAGTCAAGTGT	1372
CC	assays for receptor activity, arthritis and inflammation, leukaemias and	Qy	1343	ATTGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTGATGGCTTACAGCAACA	1402
CC	C.N.S disorders.	Db	1373	ATTGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTGATGGCTTACAGCAACA	1432
CC	Note: The sequence data for this patent did not form part of the printed	Qy	1403	GTACTCAACAGCAGCTGACTTATGTTTATTTGGACATTTGGACATTTGGACATTTGGACAT	1462
CC	specification.	Db	1433	GTACTCAACAGCAGCTGACTTATGTTTATTTGGACATTTGGACATTTGGACATTTGGACAT	1492
SQ	Sequence 1716 BP; 572 A; 297 C; 347 G; 500 T; 0 other;	Qy	1463	GAGCTCAATGAGTACAAAAGGAGTACTCTGTGACATCTTGGATGTAAACACTTGA	1522
	Query Match 76.9%; Score 1532.8; DB 22; Length 1716;	Db	1493	GAGCTCAATGAGTACAAAAGGAGTACTCTGTGACATCTTGGATGTAAACACTTGA	1552
	Best Local Similarity 99.9%; Pred. No. 5.2e-304;	Qy	1523	TTTGGTATAGAAATACCCATTTGAAATTTCTGCTGTGCGAGGTTGGTAAATTTACTTTT	1582
	Matches 1534; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	Db	1553	TTTGGTATAGAAATACCCATTTGAAATTTCTGCTGTGCGAGGTTGGTAAATTTACTTTT	1612
QY	143 TTTCAGAACACACAGAAAGGGGAGATCATTTTGAATCAGCAGATAGGTAGCAAAAT	Qy	1583	TTGGGTATATTTCTTATATATATATGATGATGATGATGATGATGATGATGATGATGATGAT	1642
DB	173 TTTCGGAACACACAGAAAGGGGAGATCATTTTGAATCAGCAGATAGGTAGCAAAAT	Db	1613	TTGGGTATATTTCTTATATATATATGATGATGATGATGATGATGATGATGATGATGATGAT	1672
QY	203 AAAGAGATATTCAAAGTACCAATAGGAGGAGTATGTTGCCCTAATATATGTTAGCT	Qy	1643	TTTTTAATAAGACTAACACAACTTAATGATTAATAA 1678	
DB	233 AAAGAGATATTCAAAGTACCAATAGGAGGAGTATGTTGCCCTAATATATGTTAGCT	Db	1673	TTTTTAATAAGACTAACACAACTTAATGATTAATAA 1708	
QY	263 GAAGGTGAAGATATTTTATGAAAGAGGTATGTTGTAATGGAAGATGATTTCCAGTG	Qy			
DB	293 GAAGGTGAAGATATTTTATGAAAGAGGTATGTTGTAATGGAAGATGATTTCCAGTG	Db			
QY	323 GAGGCTGTGACACACCTAATCTCCCGAACCCTGCTGATGGAATAGCAATTCCTCA	Qy			
DB	353 GAGGCTGTGACACACCTAATCTCCCGAACCCTGCTGATGGAATAGCAATTCCTCA	Db			
QY	383 TATGTAGATCTTTTTCAGGATCCCTCTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG	Qy			
DB	413 TATGTAGATCTTTTTCAGGATCCCTCTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG	Db			
QY	443 GTCTTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTGTGACAGAGCTGAACAT	Qy			
DB	473 GTCTTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTGTGACAGAGCTGAACAT	Db			
QY	503 TGCTCTGTCTATAGAGATATCTTGAATTTCTATGTTACTTGAATCAAACTAACAGAAAT	Qy			
DB	533 TGCTCTGTCTATAGAGATATCTTGAATTTCTATGTTACTTGAATCAAACTAACAGAAAT	Db			
QY	563 CATGGTGATTTCTCTGATGCCAGCTTCTCTTCTAGAGGATCATTTGGCCCCCAAAATAT	Qy			
DB	593 CATGGTGATTTCTCTGATGCCAGCTTCTCTTCTAGAGGATCATTTGGCCCCCAAAATAT	Db			
QY	623 GAATTTCTTAAAGTCAAGAGGGAAGAGTTTCAAGATATCTACAGAAATCTTCGACGAT	Qy			

RESULT 5

1718 GAAGGTGAAGATGATTTTATTGGAAGAGGATTTGTTGTAATCGAAGATGATTCAGTG 1777  
 323 GAGCTGTGAGCACACCTTAATCTCCCGAAACCTTGCTGCTCATGGAAAAATAGCATTTCCA 382  
 1778 GAGCTGTGAGCACACCTTAATCTCCCGAAACCTTGCTGCTCATGGAAAAATAGCATTTCCA 1837  
 383 TATGTAGACTTTTTTTGAGGATCCCTCTCTGAAAGGAGGAGAAAAAGAAAGAAATTCCT 442  
 1838 TATGTAGACTTTTTTTGAGGATCCCTCTCTGAAAGGAGGAGAAAAAGAAAGAAATTCCT 1897  
 443 GTGTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCCTGACAT 502  
 1898 GTGTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCCTGACAT 1957  
 503 TGGTCTGTCTATGAAGATATCTTGAATTCCTGAAATCTGAAATCAAAATCAACAGAAATTT 562  
 1958 TGGTCTGTCTATGAAGATATCTTGAATTCCTGAAATCTGAAATCAAAATCAACAGAAATTT 2017  
 563 CATGGTGCATTTCTCTGATGCCAGCTTCTTCTTAAAGAGGATCAATGGCCCCCAAAATTTAT 622  
 2018 CATGGTGCATTTCTCTGATGCCAGCTTCTTCTTAAAGAGGATCAATGGCCCCCAAAATTTAT 2077  
 623 GAATTTCTTAAAGTCAAAAGAGGAGAGTTCCAAAGATATCTCAGAAACTTTCTGAGCAT 682  
 2078 GAATTTCTTAAAGTCAAAAGAGGAGAGTTCCAAAGATATCTCAGAAACTTTCTGAGCAT 2137  
 683 CCAGAACTGAGTAACTAGTCAACTCTGCGAGACTTTCTTCCCTTAATGTTGGGAAACA 742  
 2138 CCAGAACTGAGTAACTAGTCAACTCTGCGAGACTTTCTTCCCTTAATGTTGGGAAACA 2197  
 743 CAATTTCTTGATAAGATACCTACAGATGTAATCTTGGGAAATTTATAAAATCTGTTCTCT 802  
 2198 CAATTTCTTGATAAGATACCTACAGATGTAATCTTGGGAAATTTATAAAATCTGTTCTCT 2257  
 803 GGAAAACTTAATGAAAGAGAGAGGTGAGCAATTTGGAACCTTTTATCATGAATTTCAATPAAT 862  
 2258 GGAAAACTTAATGAAAGAGAGAGGTGAGCAATTTGGAACCTTTTATCATGAATTTCAATPAAT 2317  
 863 TCTTGTGAGTCTCAAAAGCCTTAAACCAAGTAGACAGCACTGACCATTTCTCAGCCCTACT 922  
 2318 TCTTGTGAGTCTCAAAAGCCTTAAACCAAGTAGACAGCACTGACCATTTCTCAGCCCTACT 2377  
 923 TCAGAAACACACAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTGTGAA 982  
 2378 TCAGAAACACACAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTGTGAA 2437  
 983 AATACAGAGAGAGAGAGAGAGTCAATTTATTTATGAGAGGTGATGATCTGTAGAGGAGTC 1042  
 2438 AATACAGAGAGAGAGAGAGTCAATTTATTTATGAGAGGTGATGATCTGTAGAGGAGTC 2497  
 1043 TATGATTAACCTGATGATGATGAGGAGGATTTTCCAGGTTCTGAGTGGCTTCATCAT 1102  
 2498 TATGATTAACCTGATGATGAGGAGGATTTTCCAGGTTCTGAGTGGCTTCATCAT 2557  
 1103 CTCTTAATGGGAACCTCGAATCTCTTTAAACACCTGGAATGATGATGATGATGATGAT 1162  
 2558 CTCTTAATGGGAACCTCGAATCTCTTTAAACACCTGGAATGATGATGATGATGATGAT 2617  
 1163 CTTCAGTGTAACTAGACAGCTATTTTCAGGAGCAGCGTTTGGTCTCACTCAATCAACTT 1222  
 2618 CTTCAGTGTAACTAGACAGCTATTTTCAGGAGCAGCGTTTGGTCTCACTCAATCAACTT 2677  
 1223 CTCAGAGATGCTATATTTCTGTGAAAAACACTGAACTCGCTCTCTCCAGAGTAAGCAAAA 1282  
 2678 CTCAGAGATGCTATATTTCTGTGAAAAACACTGAACTCGCTCTCTCCAGAGTAAGCAAAA 2737  
 1283 GGACAAACACAGCTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 1342  
 2738 GGACAAACACAGCTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 2797  
 1343 ATTGTTGAAGAAACCAAGTATGAAAGCATCAGACTTTGTTTGGTGGCTTACAGCAACA 1402  
 2798 ATTGTTGAAGAAACCAAGTATGAAAGCATCAGACTTTGTTTGGTGGCTTACAGCAACA 2857

ABQ79518  
 ID ABQ79518 standard; cDNA; 3145 BP.  
 AC AC  
 XX ABQ79518;  
 DT 25-NOV-2002 (first entry)  
 XX cDNA encoding a protein similar to human sorting nexin.  
 DE Atherosclerosis; antiarteriosclerotic; marker; cardiovascular;  
 KW SSH 6; gene; human; nexin; ss.  
 XX Homo sapiens.  
 XX  
 TH Key Location/Qualifiers  
 FT CDS 128..2968  
 FT /\*tag= a  
 XX  
 XX W0200262839-A2.  
 XX  
 XX 15-AUG-2002.  
 XX  
 XX 05-FEB-2002; 2002MO-EP01327.  
 XX  
 XX 07-FEB-2001; 2001EP-0200439.  
 XX  
 XX (UYMA-) UNIV MAASTRICHT.  
 XX  
 XX Daemen WJAP, Cleutjens CBJM, Zaman GJR;  
 XX  
 XX WPI; 2002-643400/69.  
 XX  
 XX P-PSDB; ABB81193.  
 XX  
 XX Use of a polynucleotide differentially expressed in ruptured and stable  
 XX atherosclerotic plaques as a marker for atherosclerosis, useful in  
 XX treating, diagnosing or preventing atherosclerosis -  
 XX  
 XX Example 6; Page 35-37; 44pp; English.  
 XX  
 XX The invention relates to the use of a polynucleotide differentially  
 XX expressed in ruptured and stable atherosclerotic plaques as a marker  
 XX for atherosclerosis, where the polynucleotides can be selected from the  
 XX sequences shown in ABQ79517-19. The polynucleotides are useful as a  
 XX marker of atherosclerosis, which may be used: (i) in the diagnosis,  
 XX prevention and treatment of atherosclerosis; (ii) as serum/plasma markers  
 XX to screen patients at risk for plaque instability to evaluate the effects  
 XX of other treatments; (iii) in the preparation of vector molecules for the  
 XX expression of the encoded protein in host cells; and (iv) in the  
 XX identification of functional targets or analogues of the gene. The  
 XX polynucleotides, the encoded proteins or antibodies against the proteins  
 XX may be used to target other therapeutics to an unstable plaque.  
 XX Modulation of the expression of the polynucleotide can increase plaque  
 XX stability and therefore inhibit the progression of atherosclerotic  
 XX cardiovascular disease. Modulators may be used to prepare pharmaceuticals  
 XX for atherosclerotic disorders. The present sequence represents a cDNA  
 XX encoding a protein similar to human sorting nexin.  
 XX  
 XX Sequence 3145 BP; 985 A; 570 C; 659 G; 931 T; 0 other;  
 XX  
 XX Query Match 75.5%; Score 1504; DB 24; Length 3145;  
 XX Best Local Similarity 99.5%; Pred. No. 4.6e-298;  
 XX Matches 1529; Conservative 0; Mismatches 5; Indels 2; Gaps 2;  
 XX  
 XX 143 TTACGAGACACACAGAAAGGGGAGATCTTTGGATCAGCAGATAGTACCAAAAT 202  
 XX 1598 TTTCCGAAACACACAGAAAGGGGAGATCTTTGGATCAGCAGATAGTACCAAAAT 1657  
 XX  
 XX 203 AAAGGAGTATTCAAAGTACCACAAATGGAGGAGCTATGTTGCTAATATGTTAGTGT 262  
 XX 1658 AAAGGAGTATTCAAAGTACCACAAATGGAGGAGCTATGTTGCTAATATGTTAGTGT 1717  
 XX  
 XX 263 GAAGGTGAAGATGATTTTATTGAAAGAGGATTTGTTGTAATGGAAGATGATTTCCAGTG 322



953 TCAGAAACCAACAGAGCTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAA 1012  
983 AATACAGAGAGAGCAACCAATCAGATTTATTTATGGAGGTGATGACTGTAGAGGAGTC 1042  
1013 AATACAGAGAGAGCAACCAATCAGATTTATTTATGGAGGTGATGACTGTAGAGGAGTC 1072  
1043 TATGATTACCTGATGATGATGAGCGGTAGTTTTCCAGGTTCTCTGACTGGCTTCATCAT 1102  
1073 TATGATTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1092  
1103 CTCCTTAATGGGAACCTGGAATCCTCTTTTAAACACCCCTGGAAATGATGATGATGATGAT 1162  
1093 -----AGATGCTATATCTCTGGAACACTGAACTCGCTCTCTCCAGATAAGCAAAA 1092  
1163 CTTGAGTGAATAGACAGCTATTTTCCAGGACCGGTTTGGTCTCTACTATACACTT 1222  
1093 ----- 1092  
1223 CTCAGAGATGCTATATCTCTGGAACACTGAACTCGCTCTCTCCAGATAAGCAAAA 1282  
1093 -----AGATGCTATATCTCTGGAACACTGAACTCGCTCTCTCCAGATAAGCAAAA 1147  
1283 GGAGCAAAACAGACCTTTTGAAGAAATGATGAATTAATCCAGATCTGTTAGTCAAGTGT 1342  
1148 GGAGCAAAACAGACCTTTTGAAGAAATGATGAATTAATCCAGATCTGTTAGTCAAGTGT 1207  
1343 ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCAACA 1402  
1208 ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCAACA 1267  
1403 GTACTCAACAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1462  
1268 GTACTCAACAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1327  
1463 GAGCTCAATGAAGTACAAAGGAAGTACTCTCTGACATCTGGAATGAAACACTTGA 1522  
1328 GAGCTCAATGAAGTACAAAGGAAGTACTCTCTGACATCTGGAATGAAACACTTGA 1387  
1523 TTGGTATAGATTAACCCATTTGAAATTTCTGCTGTCGAGGGTGGTGAATTTACTTTT 1582  
1388 TTGGTATAGATTAACCCATTTGAAATTTCTGCTGTCGAGGGTGGTGAATTTACTTTT 1447  
1583 TTGGTATAGATTAACCCATTTGAAATTTCTGCTGTCGAGGGTGGTGAATTTACTTTT 1642  
1448 TTGGTATAGATTAACCCATTTGAAATTTCTGCTGTCGAGGGTGGTGAATTTACTTTT 1507  
1643 TTTTATAGAGCTAACACAACTTAATGATTAATA 1678  
1508 TTTTATAGAGCTAACACAACTTAATGATTAATA 1543

RESULT 7  
AAZ17379  
ID AAZ17379 standard; cDNA; 779 BP.  
XX AC AAZ17379;  
XX AC AAZ17379;  
DT 12-OCT-1999 (first entry)  
XX DE Human gene expression product cDNA sequence SEQ ID NO:4852.  
XX DE Human; gene; gene expression product; diagnosis; therapy; probe;  
KW detection; mapping; tissue typing; profiling; forensic; cancer;  
KW Genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
XX OS Homo sapiens.  
XX OS  
PN WO9938972-A2.  
XX PD  
XX PD 05-AUG-1999.  
XX PD  
XX PD 28-JAN-1999; 99WO-US01619.

XX PR 03-APR-1998; 98US-0080666.  
PR 28-JAN-1998; 98US-0072910.  
PR 24-FEB-1998; 98US-0075954.  
PR 31-MAR-1998; 98US-0080114.  
PR 03-APR-1998; 98US-0080515.  
XX (CHIR) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
XX Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;  
PI Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;  
PI Jones WL, Kassam A, Kennedy GC, Kita D, Labat I,  
PI Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;  
PI Stache-Crain B, Sudduth-Klinger J, Williams LT;  
XX WPI; 1999-494092/41.  
XX Novel human genes and their expression products which are  
PT differentially expressed in different cell types  
PT Claim 1; Page 2307; 2479pp; English.  
XX The present invention describes a library of human polynucleotides  
CC comprising the sequences given in AAZ12532 to AAZ17779. Also described is  
CC a method of detecting differentially expressed genes correlated with the  
CC cancerous state of a mammalian cell, comprising detecting at least one  
CC differentially expressed gene product in a test sample from a cell  
CC suspected of being cancerous, where the gene product is encoded by one  
CC of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The  
CC polynucleotides can be used as a source of primers and probes, which can  
CC be used for a variety of purpose, e.g. detection of expression levels,  
CC mapping, tissue typing or profiling, forensics, genetic analysis and  
CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
CC can be used for raising antibodies for experimental, diagnostic and  
CC therapeutic purposes. The polynucleotides may also be used to construct  
CC arrays for diagnostics (which may be used to determine function of an  
CC encoded protein); and to detect differences in expression levels between  
CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
CC identify a genetic predisposition or susceptibility to a disease such as  
CC cancer). The polynucleotides of the invention are especially used in the  
CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
CC and lung cancer. The polynucleotides can also be used to screen for  
CC peptide analogues and antagonists.  
XX Sequence 779 BP; 248 A; 143 C; 153 G; 232 T; 3 other;  
SQ Query Match 32.9%; Score 655.4; DB 20; Length 779;  
Best Local Similarity 98.3%; Pred. No. 1.2e-124;  
Matches 704; Conservative 0; Mismatches 7; Indels 5; Gaps 4;  
QY 882 CTAAACCAAGTAGACCGAG-AACTGACCATTTCTGAGCCCTACTTCAGAAACCAACAGAG 940  
DB 59 CGAGACCAAGTAGACCGAGAACTGACCATTTCTGAGCCCTACTTCAGAAACCAACAGAG 118  
QY 941 CTTTTCATGATCTGTTTAAATAATGCAACCGTCTGAAATATACAGAGAGAAAGCAA 1000  
DB 119 CTTTTCATGATCTGTTTAAATAATGCAACCGTCTGAAATATACAGAGAGAAAGCAA 178  
QY 1001 AATCAGAAATATTTATGGAGGTGATGACTGTAGAGAGGTCTATGATTAATCTGATGAT 1060  
DB 179 AATCAGAAATATTTATGGAGGTGATGACTGTAGAGAGGTCTATGATTAATCTGATGAT 238  
QY 1061 GTAGACCGGTAGTTTCCAGGTTCCGAGTCTGATGCTTATCTCTTAATGGGAACCTGA 1120  
DB 239 GTAGACCGGTAGTTTCCAGGTTCCGAGTCTGATGCTTATCTCTTAATGGGAACCTGA 298  
QY 1121 ATCTCTTTTAAACACACCCCTGGAATGATGATTAATCTCTGAGTGAACCTAGAA 1180  
DB 299 ATCTCTTTTAAACACACCCCTGGAATGATGATTAATCTCTGAGTGAACCTAGAA 358  
QY 1181 CAGCTATTTCAGGAGCACCCTGTTGCTCTCACTATAACACTCTCAGAGATGCTATATTC 1240

359	DB	CAGCTATTTCAGGAGCACCGTTTGGCTCTCACTCATACACTTCTCCAGAGATGCTATATTC	418
1241	QY	TGTGAAAACACTGAACCTCGCTCTCTCCAGAGATAGCAAAAAGGAGCAAAAACAGACTTTT	1300
419	DB	TGTGAAAACACTGAACCTCGCTCTCTCCAGATAGCAAAAAGGAGCAAAAACAGACTTTT	478
1301	QY	GAAGAAATGATGAATTACATTCAGATCTGTTAGTCAGTGTATTGGTGAAGAAACCAAG	1360
479	DB	GAAGAAATGATGAATTACATTCAGATCTGTTAGTCAGTGTATTGGTGAAGAAACCAAG	538
1361	QY	TATGAAAGCATCAGACATTCCTGTTTGATGGCTTACAGCAACCACTACTCAAACAAGCAGCTG	1420
539	DB	TATGAAAGCATCAGACATTCCTGTTTGATGGCTTACAGCAACCACTACTCAAACAGCAGCTG	598
1421	QY	ACTTATGTTTTATTGGACATTTGTGATACAGAACTCTTTCCAGAGCTCAATAAGGTACAA	1480
599	DB	ACTTATGTTTTATTGGACATTTGTGATACAGAACTCTTTNCAGAGCTCAATAAGGTACAA	658
1481	QY	AAGGAAGTTACTCTGTGCATCTT--GGATGTAAACACTTT--GGATTGGTATAGATAA	1537
659	DB	AAGGAAGTTACTCTGTGCATCTTGGGATGTAAACACTTTGGGATTTGGTATAGATAA	718
1538	QY	CCCAATTGAAATTTCTGCTGTG--CGAGGGTGGTGAATAATTTACTTTTTTGGGTATAT	1592
719	DB	CCCAATTGAAATTTCTGCTGTGCGAAGGGTGGTGAATAATTTACTTTTTTGGGTATAT	774

## RESULT 8

ABE63556/C  
ID ABE63556 standard; DNA; 451 BP.  
XX  
XX ABE63556;  
XX  
XX 15-MAY-2002 (first entry)  
XX  
XX Breast cancer related gene sequence SEQ ID NO:1893.  
XX  
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
XX stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200194629-A2.  
XX  
XX 13-DEC-2001.  
XX  
XX 30-MAY-2001; 2001WO-US10838.  
XX  
XX 05-JUN-2000; 2000US-209473P.  
PR  
XX 05-JUN-2000; 2000US-209531P.  
PR  
XX 18-SEP-2000; 2000US-233133P.  
PR  
XX 18-SEP-2000; 2000US-233617P.  
PR  
XX 20-SEP-2000; 2000US-234009P.  
PR  
XX 20-SEP-2000; 2000US-234034P.  
PR  
XX 20-SEP-2000; 2000US-234052P.  
PR  
XX 22-SEP-2000; 2000US-234509P.  
PR  
XX 22-SEP-2000; 2000US-234567P.  
PR  
XX 25-SEP-2000; 2000US-234923P.  
PR  
XX 25-SEP-2000; 2000US-234924P.  
PR  
XX 25-SEP-2000; 2000US-235077P.  
PR  
XX 25-SEP-2000; 2000US-235082P.  
PR  
XX 25-SEP-2000; 2000US-235134P.  
PR  
XX 25-SEP-2000; 2000US-235280P.  
PR  
XX 26-SEP-2000; 2000US-235637P.  
PR  
XX 26-SEP-2000; 2000US-235638P.  
PR  
XX 27-SEP-2000; 2000US-235711P.  
PR  
XX 27-SEP-2000; 2000US-235720P.  
PR  
XX 27-SEP-2000; 2000US-235840P.  
PR  
XX 27-SEP-2000; 2000US-235863P.  
PR  
XX 28-SEP-2000; 2000US-236028P.  
PR  
XX 28-SEP-2000; 2000US-236032P.  
PR



1777 AATGATGTTATGCAACAGACAGTAAGACTGATTAACCTGCTATTTGTATAGCTTTGAAATA 1836  
205 AATGATGTTATGCAACAGACAGTAAGACTGATTAACCTGCTATTTGTATAGCTTTGAAATA 146  
1837 AATGATGTTATGCAACAGACAGTAAGACTGATTAACCTGCTATTTGTATAGCTTTGAAATA 1896  
145 AATGATGTTATGCAACAGACAGTAAGACTGATTAACCTGCTATTTGTATAGCTTTGAAATA 86  
1897 AATGATGTTATGCAACAGACAGTAAGACTGATTAACCTGCTATTTGTATAGCTTTGAAATA 1956  
85 AATGATGTTATGCAACAGACAGTAAGACTGATTAACCTGCTATTTGTATAGCTTTGAAATA 26  
1957 GCATCCCTATATGAGTATTTTAAAT 1981  
25 GCATCCCTATATGAGTATTTTAAAT 1

RESULT 9  
ABS77486  
ID ABS77486 standard; cDNA; 725 BP.  
AC ABS77486;  
DT 12-DEC-2002 (first entry)  
DE Prog embryonic gene sequence Q9925894.  
KW Prog; ss; embryonic development; developmental disorder;  
KW microarray; cell differentiation.  
OS Xenopus laevis.  
XX US2002081610-A1.  
XX 27-JUN-2002.  
XX 23-JUL-2001; 2001US-0910943.  
XX 21-JUL-2000; 2000US-219658P.  
XX (UVRQ) UNIV ROCKFELLER.  
XX Hemmati-Brivanlou A, Altmann CR;  
XX WPI; 2002-626534/67.  
XX Nucleic acid array containing Xenopus embryonic nucleic acids is useful  
XX to identify genes involved in embryonic development, to identify  
XX different types of embryonic cells, and to diagnose developmental  
XX disorders

Claim 1; Page 756; 823pp; English.  
The invention relates to a nucleic acid array, where each coordinate  
contains a single nucleic acid species having one of 770 nucleotide  
sequences (appearing as ABS76747-ABS77516) a of a Xenopus embryonic gene  
product, or its complement or hybridisable fragment of not less than  
20 contiguous nucleotides of one of those sequences. Also included are  
detecting differential expression of embryonic genes, comprising:  
(a) contacting a nucleic acid array comprising genes expressed in  
embryonic but not mature cells with nucleic acids from sample and  
control cells; and (b) detecting differential hybridisation of nucleic  
acids from the sample cells relative to the control cells; and detecting  
defects in development, comprising: (a) contacting nucleic acids from  
test cells undergoing development with a nucleic acid array of gene  
products known to play a fundamental role in the development process; and  
(b) detecting a difference in expression of a fundamental gene in the  
sample cells relative to a standard. The invention is useful to identify  
genes involved in embryonic development and related processes such as  
cell differentiation. This would be useful for diagnosing developmental  
disorders and for identifying different types of embryonic cells.  
The present sequence is one of the 770 Xenopus embryonic cDNA sequences.

SQ Sequence 725 BP; 229 A; 135 C; 140 G; 218 T; 3 other;  
Query Match 21.3%; Score 424.4; DB 24; Length 725;  
Best Local Similarity 77.7%; Pred. No. 2e-77;  
Matches 512; Conservative 0; Mismatches 147; Indels 0; Gaps 0;  
QY 731 GGTGGGAAACACAAATTTCTTGAATAGATACCTACCATGTAATCTTTGGGAAATATATA 790  
DB 62 GGTGGGAAATCTAGTTTCATGCAAAATCTTCCAGATGTGACCTTGGGAAATATATC 121  
QY 791 AAATCTGTCTGCTGAAACAACTAATGAAAGAGAAAGGTCTAGCATTTGGAACTTTTATCATG 850  
DB 122 AAGTCAGTTCTGGAATCACTTAAAGAAAGGCAACATCTGGAGCTGTTTATAATG 181  
QY 851 AATTTCAATTAATCTTGTGAGTCTTCCAAAGCCTTAACCAAGTAGACAGACTGACCATTT 910  
DB 182 AATTTCAATTAATCTTGTGAGTCTTCCAAAGCCTTAAAGCCTTAAAGCCTTAAAGCCTT 241  
QY 911 CTCAGCCCTCTTCCAGAAACCAACAGAGCTTTTCAATGATCTGTTTAAATAATAATGCA 970  
DB 242 CTAAGCCCGACTTCCAGAAACCAATAAAGCTTTTAAATGATCTGTAAGATTAATGCT 301  
QY 971 AACCGTCTGCTGAAATACAGAGAGAAAGCAAAATCAGAAATTTATTTATGGAGGTGATGACT 1030  
DB 302 AATCGCTCTGAAATACAGAGAGAAAGCAATTAACAGAAATTTATTTATGGAGGTGATGACT 361  
QY 1031 GTAGAAGAGTCTATGATTAACCTGATGTATGAGACGGGTAGTTTTCAGAGTTCCTGAC 1090  
DB 362 GTAGAAGGTGTTTATGACTATTTAATGTATATAGGAGGTGTTGTTTTCATTTCTGAT 421  
QY 1091 TGGCTTCTATCATCTCTTAATGGGAATCTGAAATCTCTTTAAACACCCCTGGAATGAT 1150  
DB 422 TGGTTTCTATCATCTCTTTTGAAGGTGGAGCAATCTCTTTAAACATACCCCTGAGACTTAC 481  
QY 1151 ACTGATTTACTATCTTCTAGTGTAAACTAGAACAGCTATTTTCAGAGACCCGTTTCGTCTCA 1210  
DB 482 ACAAATAGCTATTTAACTATTAATTAAGAGCGCTCTTTCAAGAGCATGCTGTTCT 541  
QY 1211 CTCATAACACTTCTCAGAGATCTATATTTCTGTGAAACACTGAACTTCTCTCTCCAA 1270  
DB 542 TTGATTAATCTGCTTAGAGTGTATATTTTGTGAAACCCGCTGAACCAAGCATATTACAT 601  
QY 1271 GATAAGCAAAAGGAGCAAAAACAGACTTTTGAAGAAATGATGAATTAATCCAGATCTG 1330  
DB 602 CTGAAGCAGCAAAAGGCAAAAGCTTACATTTGAAGAAATGATGCGCTATATTCCAGATTG 661  
QY 1331 TTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGG 1389  
DB 662 ATTGTTAAATGTATTTGGTGTATGAAGCAATTAATGAAGCATTCGACTTCTGTTGATGG 720

RESULT 10  
ABX50611  
ID ABX50611 standard; cDNA; 402 BP.  
XX AC ABX50611;  
XX DT 25-FEB-2003 (first entry)  
XX DE Bovine EST associated with lactation/muscle/fat deposition #540.  
XX KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
XX muscle deposition; fat deposition; genome mapping; gene identification;  
XX Gene analysis; cattle breeding.  
XX OS Bos Taurus.  
XX PN US2002137160-A1.  
XX 26-SEP-2002.  
XX 26-OCT-2001; 2001US-0983965.  
XX

PR 17-DEC-1998; 98US-113678P.  
PR 15-DEC-1999; 99US-0465231.  
XX  
PA (BYAT//) BYATT J C.  
PA (MATH//) MATHIALAGAN N.  
PA (TAON//) TAO N.  
PA (WARR//) WARREN W C.  
XX  
PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX  
XX WPI; 2003-102386/09.  
XX  
XX Purified nucleic acid molecules, useful for genome mapping, gene  
PT identification and analysis, cattle breeding or preparation of  
PT constructs for cattle gene expression and genetically improved cattle -  
XX  
XX  
PS Claim 2; SEQ ID No 540; 38pp; English.  
XX  
XX The invention relates to a purified nucleic acid molecule associated with  
XX lactation or muscle and fat deposition (designated LMFD), derived  
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a  
CC second nucleic acid molecule comprising any of 5912 nucleotide  
CC sequences, appearing as ABX50072-ABX55983, or complements of them.  
CC Also included are; (1) a transformed cell having a nucleic acid  
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
CC nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridisation between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the complementary  
CC nucleic acid is predictive of the level or pattern of the molecule.  
CC The LMFD nucleic acid is used for determining a level or pattern  
CC of a molecule in a bovine cell or tissue. It is useful for genome  
CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 5912 bovine  
CC LMFD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?DocID=20020137160.  
XX  
XX Sequence 402 BP; 141 A; 71 C; 76 G; 113 T; 1 other;  
XX  
XX  
XX Query Match 18.3%; Score 384.2; DB 25; Length 402;  
XX Best Local Similarity 94.0%; Pred. No. 3.6e-65;  
XX Matches 378; Conservative 0; Mismatches 24; Indels 0; Gaps 0;  
XX  
QY 726 CTAATGGTGGGAAACACAAATTTCTTGATAAGATACTACAGATGTAATCTTGGGAAA 785  
DB 1 CCATGGTGGGAAACACAGTTCTTGATAAGATACTACAGATGTAATCTTGGGAAA 60  
QY 786 TTATAAATCTGTTCTGAAACAACTAATGAAGAGAAAGGTGAGCATTTGAACTTTTA 845  
DB 61 TTATAAATCTGTTCTGAAACAACTAATGAAGAGAAAGGTGAGCATTTGAACTTTCA 120  
QY 846 TCATGAATTTCAATTAATTTCTGTGAGTCTCCAGCCCTAAACCAAGTAGACCACTGA 905  
DB 121 TCATGAATTTCAATTAATTTCTGTGAGTCTCCAGCCCTAAACCAAGTAGACCACTGA 180  
QY 906 CCATTTCTGAGCCCTACTCTCAGAAACACACAGAGCTTTTCAATGATCTGTTAAATA 965  
DB 181 CCATTTCTGAGCCCTACTCTCAGAGATAATAGAGAGCTTTTCAATGATCTGTTAAATA 240  
QY 966 ATGCAAAACCGTCTGCAAAATACAGAGAGAGCAAAATCAGAAATTTATGAGAGTGA 1025  
DB 241 ATGCAAAACCGTCTGCAAAATACAGAGAGAGCAAAATCAGAAATTTATGAGAGTGA 300  
QY 1026 TGACTGTAGAGAGGTCTTATGANTTACCTGATGTATGTAGGACGGGTAGTTTCCAGGTT 1085

DB 301 TGACTGTAGAGAGGTCTTATGATTTACCTGATGTATGTANGACAGGTGTTTTCCAGTTC 360  
QY 1086 CTGACTGCTTCATCATCTCTTAATGGAATCGAATCTCT 1127  
DB 361 CTGACTGCTTCATCATCTCTTAATGGAATCGAATCTCT 402  
RESULT 11  
AAC02755  
ID AAC02755 standard; cDNA; 358 BP.  
XX  
XX AAC02755;  
XX AC  
XX DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein 5' EST, SEQ ID NO: 2753.  
XX  
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
XX gene therapy; chromosome mapping; ss.  
XX Homo sapiens.  
XX  
XX EP1033401-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 21-FEB-2000; 2000EP-0200610.  
XX  
XX 26-FEB-1999; 99US-0122487.  
XX  
XX (GEST ) GENSET.  
XX  
XX Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
XX WPI; 2000-500381/45.  
XX  
XX P-PSDB; AAG02749.  
XX  
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 2753; 71pp + CD-ROM; English.  
XX  
XX The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
XX  
XX Sequence 358 BP; 125 A; 63 C; 69 G; 101 T; 0 other;  
XX  
XX Query Match 17.9%; Score 357; DB 21; Length 358;  
XX Best Local Similarity 100.0%; Pred. No. 1e-63;  
XX Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 760 ACTACAGATGTAATCTTGGGAAATATATAAATCTCTTCTCGAAACATTAAGAAGA 819  
DB 1 ACTACAGATGTAATCTTGGGAAATATATAAATCTCTTCTCGAAACATTAAGAAGA 60  
QY 820 GAAAGTCAGCATTTGGAACTTTTATCATGATTAATTTCTTGTGAGTCTCCAAA 879  
DB 61 GAAAGTCAGCATTTGGAACTTTTATCATGATTAATTTCTTGTGAGTCTCCAAA 120

QY 880 GCTAAACCAAGTAGACCAAGTACCAATTCAGCCCTTCTCAGAAACCAACAGAA 939  
 Db 121 GCTAAACCAAGTAGACCAAGTACCAATTCAGCCCTTCTCAGAAACCAACAGAA 180  
 QY 940 GCTTTCAATGATCTGTTTAAATAATGCAACCGTGTGAAATACAGAGGAAGCA 999  
 Db 181 GCTTTCAATGATCTGTTTAAATAATGCAACCGTGTGAAATACAGAGGAAGCA 240  
 QY 1000 AATACAGAAATATTTATGAGGTGATGACTGAGAGGAGTCTATGATTACTGATGTA 1059  
 Db 241 AATACAGAAATATTTATGAGGTGATGACTGAGAGGAGTCTATGATTACTGATGTA 300  
 QY 1060 TGTAGGAGGGTAGTTTCCAGGTTCTGACTGGCTTCATCTCTTAATGGGAAC 1116  
 Db 301 TGTAGGAGGGTAGTTTCCAGGTTCTGACTGGCTTCATCTCTTAATGGGAAC 357

## RESULT 12

ABX50426  
 ID ABX50426 standard; cDNA; 305 BP.

AC ABX50426;

XX 25-FEB-2003 (first entry)

XX Bovine EST associated with lactation/muscle/fat deposition #355.

XX Bovine, ss; EST, expressed sequence tag; lactation; LMFD;  
 XX muscle deposition; fat deposition; genome mapping; gene identification;  
 XX gene analysis; cattle breeding.

OS Bos Taurus.

XX US2002137160-A1.

XX 26-SEP-2002.

XX 26-OCT-2001; 2001US-0983965.

XX 17-DEC-1998; 98US-113678P.

XX 15-DEC-1999; 99US-0465231.

XX (BYATT/) BYATT J C.

XX (MATH/) MATHIALAGAN N.

XX (TAON/) TAO N.

XX (WARR/) WARREN W C.

PI Byatt JC, Mathialagan N, Tao N, Warren WC;

XX WPI; 2003-102386/09.

XX Purified nucleic acid molecules, useful for genome mapping, gene  
 XX identification and analysis, cattle breeding or preparation of  
 XX constructs for cattle gene expression and genetically improved cattle -

PS Claim 2; SEQ ID No 355; 38pp; English.

XX The invention relates to a purified nucleic acid molecule associated with  
 XX lactation or muscle and fat deposition (designated LMFD), derived  
 XX from cattle, and the LMFD nucleic acid can specifically hybridize to a  
 XX second nucleic acid molecule comprising any of 5912 nucleotide  
 XX sequences, appearing as ABX50072-ABX55983, or complements of them.  
 XX Also included are; (1) a transformed cell having a nucleic acid  
 XX comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
 XX translated sequence that functions in the cell to cause termination of  
 XX transcription and addition of polyadenylated ribonucleotides to a 3' end  
 XX of the mRNA molecule; and (2) determining a level or pattern of a  
 XX molecule in a bovine cell or tissue comprising: (a) incubating a marker  
 XX nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
 XX complement or fragment) with a complementary nucleic acid molecule  
 XX obtained from the bovine cell or tissue, where hybridisation between the  
 XX marker nucleic acid and the complementary nucleic acid permits the  
 XX detection of the molecule; and (b) detecting the level or pattern of the

CC complementary nucleic acid, where the detection of the complementary  
 CC nucleic acid is predictive of the level or pattern of the molecule.  
 CC The LMFD nucleic acid is used for determining a level or pattern  
 CC of a molecule in a bovine cell or tissue. It is useful for genome  
 CC mapping, gene identification and analysis, cattle breeding, preparation  
 CC of constructs for use in cattle gene expression, or for genetically  
 CC improving cattle. The present sequence is one of the 5912 bovine  
 CC LMFD EST (expressed sequence tag) nucleic acids.

CC Note: The present sequence was not shown in the specification but  
 CC was obtained in electronic format from the USPTO web site:  
 CC seqdata.uspto.gov/sequence.html?DocID=20020137160.

XX

SQ Sequence 305 BP; 109 A; 60 C; 51 G; 85 T; 0 other;

Query Match 13.4%; Score 267.2; DB 25; Length 305;

Best Local Similarity 92.4%; Pred. No. 2.3e-45;

Matches 281; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 653 CAAGAATATCTACAGAACTTCTGCAGCATCCGAACTGAGTATAGTCAACTTCTGGCA 712

Db 1 CAGGAATATCTGCAGAACTTCTGCAACATCCAGATTGAGTAACAGTCAACTTCTGGCA 60

QY 713 GACTTTCTTCCCTTAATGTTGGGAAACACAAATTTCTTGATAGATACACAGATGTA 772

Db 61 GATTTCTTCTCCCAATGTTGGGAAACACAGTTTCTTGATAGATACACAGATGTA 120

QY 773 AATCTTGGGAAATTAATAAATCTGTTCTCGGAAACTTAATGAAGAGAAAGTTCAGCAT 832

Db 121 AATCTTGGGAAATTAATAAATCTGTTCTCGGAAACTTAATGAAGAGAAAGTTCAGCAT 180

QY 833 TTGGAACCTTTTATCATGAATTTCAATTAATTTCTGTGAGTCTCCAAAGCCTTAAACCAAGT 892

Db 181 TTGGAACCTTTTATCATGAATTTCAATTAATTTCTGTGAGTCTCCAAAGCCTTAAACCAAGT 240

QY 893 AGACCAGAACTGACCATTTCTCAGCCCTTCTCAGAAACCAACAGAAAGCTTTTCAATGAT 952

Db 241 AAACAGAACTGACCAATCTCAGCCCTTCTCAGAAACCAACAGAAAGCTTTTCAATGAT 300

QY 953 CTGT 956

Db 301 CTGT 304

## RESULT 13

AAH69552  
 ID AAH69552 standard; cDNA; 311 BP.

XX AC AAH69552;

XX 19-SEP-2001 (first entry)

XX Human cervical cancer marker nucleic acid 826.

XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

OS Homo sapiens.

XX WO200142467-A2.

XX 14-JUN-2001.

XX 08-DEC-2000; 2000WO-US33312.

XX 08-DEC-1999; 99US-0169681.

XX 21-DEC-1999; 99US-0171350.

XX 14-MAR-2000; 2000US-0189315.

XX 12-MAY-2000; 2000US-0203791.

XX 09-JUN-2000; 2000US-0210600.

XX 21-JUL-2000; 2000US-0220114.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Deeds J, Berger A, Zhao X;

XX WPI; 2001-375006/39.  
DR New isolated nucleic acid for diagnosing and treating cervical cancer  
XX and for assessing and detecting compounds for treating the cancer -  
PT  
XX  
PS Claim 1; Page 243; 1051pp; English.  
XX  
CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful; to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a pre-malignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.  
XX  
SQ Sequence 311 BP; 84 A; 67 C; 63 G; 94 T; 3 other;  
  
Query Match 4.3%; Score 86; DB 22; Length 311;  
Best Local Similarity 90.2%; Pred. No. 2.7e-08;  
Matches 92; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
  
QY 1403 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAAGTCTTTCCA 1462  
DB 24 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAAGTCTTTCCA 83  
  
QY 1463 GAGCTCAATAAGGTACAAAGGAAGTCTTACCTCTGTGACATCT 1504  
DB 84 GAGCTCAATAAGGTACGGGGAAAGTCAACTCAGTTACTCT 125  
  
RESULT 14  
AAH70887  
ID AAH70887 standard; cDNA; 313 BP.  
XX  
AC AAH70887;  
XX  
DT 19-SEP-2001 (first entry)  
XX  
DE Human cervical cancer marker nucleic acid 2161.  
XX  
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200142467-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 08-DEC-2000; 2000WO-US33312.  
XX  
PR 08-DEC-1999; 99US-0169681.  
PR 21-DEC-1999; 99US-0171350.  
PR 14-MAR-2000; 2000US-0189315.  
PR 12-MAY-2000; 2000US-0203791.  
PR 09-JUN-2000; 2000US-0210600.  
PR 21-JUL-2000; 2000US-0220114.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Deeds J, Berger A, Zhao X;  
XX  
DR WPI; 2001-375006/39.  
XX  
XX New isolated nucleic acid for diagnosing and treating cervical cancer  
XX and for assessing and detecting compounds for treating the cancer -  
XX  
PS Claim 1; Page 456; 1051pp; English.  
XX  
CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful; to assess if a patient is afflicted with

CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a pre-malignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.  
XX  
SQ Sequence 313 BP; 83 A; 70 C; 63 G; 97 T; 0 other;  
  
Query Match 4.3%; Score 86; DB 22; Length 313;  
Best Local Similarity 90.2%; Pred. No. 2.7e-08;  
Matches 92; Conservative 0; Mismatches 10; Indels 0; Gaps 0;  
  
QY 1403 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAAGTCTTTCCA 1462  
DB 2 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAAGTCTTTCCA 61  
  
QY 1463 GAGCTCAATAAGGTACAAAGGAAGTCTTACCTCTGTGACATCT 1504  
DB 62 GAGCTCAATAAGGTACGGGGAAAGTCAACTCAGTTACTCT 103  
  
RESULT 15  
AAH72445  
ID AAH72445 standard; cDNA; 299 BP.  
XX  
AC AAH72445;  
XX  
DT 19-SEP-2001 (first entry)  
XX  
DE Human cervical cancer marker nucleic acid 3719.  
XX  
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200142467-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 08-DEC-2000; 2000WO-US33312.  
XX  
PR 08-DEC-1999; 99US-0169681.  
PR 21-DEC-1999; 99US-0171350.  
PR 14-MAR-2000; 2000US-0189315.  
PR 12-MAY-2000; 2000US-0203791.  
PR 09-JUN-2000; 2000US-0210600.  
PR 21-JUL-2000; 2000US-0220114.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Deeds J, Berger A, Zhao X;  
XX  
DR WPI; 2001-375006/39.  
XX  
XX New isolated nucleic acid for diagnosing and treating cervical cancer  
XX and for assessing and detecting compounds for treating the cancer -  
XX  
PS Claim 1; Page 696; 1051pp; English.  
XX  
XX The invention relates to novel genes (AAH68727-AAH73383) associated with  
XX cervical cancer with cytostatic activity. The nucleic acids and encoded  
XX polypeptides are useful; to assess if a patient is afflicted with  
XX cervical cancer or has a pre-malignant condition; to monitor the  
XX progression of cervical cancer or a pre-malignant condition in a patient;  
XX and to select and/or assess the efficacy of a compound or therapy for  
XX inhibiting cervical cancer in a patient. The nucleic acids may also be  
XX useful for gene therapy.  
XX  
SQ Sequence 299 BP; 82 A; 64 C; 57 G; 96 T; 0 other;  
  
Query Match 4.2%; Score 84; DB 22; Length 299;  
Best Local Similarity 90.0%; Pred. No. 6.8e-08;  
Matches 90; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY	1405	ACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACCTGTTTCCAGA	1464
Db	1	ACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACCTGTTTCCAGA	60
QY	1465	GCTCAATAAGGTACAAAGGAAGTTACCTCTGTGACATCT	1504
Db	61	GCTCAATAAGGTACGGGGAAAGTCAACTCAGTTACCTCT	100

Search completed: January 31, 2004, 10:41:08  
Job time : 553 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 18:21:00 ; Search time 708 Seconds  
(without alignments)  
10254.176 Million cell updates/sec

Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: Gtatgaactccaaagt.....tttaataataaaaaaaaaa 1992

Scoring table: OLIGO\_NUC  
Gapop 60.0, Gapext 60.0

Searched: 2434938 seqs, 1822278265 residues

Word size: 6

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications NA:  
1: /cgn2\_6/prodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/prodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/prodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/prodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/prodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/prodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/prodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/prodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq:  
10: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq:  
11: /cgn2\_6/prodata/2/pubpna/US09\_PUBCOMB.seq:  
12: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/prodata/2/pubpna/US09\_NEW\_PUB.seq:  
14: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:  
15: /cgn2\_6/prodata/2/pubpna/US10\_PUBCOMB.seq:  
16: /cgn2\_6/prodata/2/pubpna/US10\_NEW\_PUB.seq:  
17: /cgn2\_6/prodata/2/pubpna/US60\_NEW\_PUB.seq:  
18: /cgn2\_6/prodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1824	91.6	3470	US-10-198-846-13016	Sequence 13016, A
2	1532	76.9	1716	US-10-117-722-922	Sequence 922, App
3	1532	76.9	1716	US-10-037-270-922	Sequence 922, App
4	918	46.1	1551	US-10-117-722-923	Sequence 923, App
5	918	46.1	1551	US-10-037-270-923	Sequence 923, App
6	444	22.3	535	US-09-918-995-22530	Sequence 22530, A
7	419	21.0	508	US-09-918-995-20205	Sequence 20205, A
8	397	19.9	397	US-09-918-995-5183	Sequence 5183, App
9	385	19.3	451	US-09-954-531-828	Sequence 828, App
10	360	18.1	471	US-10-242-535A-25415	Sequence 25415, A
11	342	17.2	446	US-10-242-535A-49594	Sequence 49594, A
12	98	4.9	464	US-09-918-995-17035	Sequence 17035, A
13	60	3.0	60	US-09-908-975-12718	Sequence 12718, A
14	59	3.0	305	US-09-983-965-355	Sequence 355, App
15	59	3.0	402	US-09-983-965-540	Sequence 540, App

16	34	1.7	384	10	US-09-783-590-4845	Sequence 4845, Ap
17	32	1.6	280	10	US-09-783-590-4866	Sequence 4866, Ap
18	23	1.2	3673778	13	US-10-312-841-1	Sequence 1, Appli
19	22	1.1	756	13	US-10-027-632-164666	Sequence 164666,
20	22	1.1	756	14	US-10-027-632-164666	Sequence 164666,
21	22	1.1	3126	12	US-10-108-260A-777	Sequence 777, App
22	22	1.1	3448	13	US-10-027-632-115054	Sequence 115054,
23	22	1.1	5807	13	US-10-311-455-1128	Sequence 1128, Ap
24	22	1.1	6009	13	US-10-311-455-802	Sequence 802, App
25	22	1.1	6070	13	US-10-311-455-1651	Sequence 1651, Ap
26	22	1.1	6070	13	US-10-240-485-131	Sequence 131, App
27	22	1.1	7049	13	US-10-311-455-130	Sequence 130, App
28	22	1.1	7049	13	US-10-240-452-6	Sequence 6, Appli
29	22	1.1	7238	13	US-10-311-455-422	Sequence 422, App
30	22	1.1	7771	13	US-10-311-455-1946	Sequence 1946, Ap
31	22	1.1	8576	13	US-10-311-455-2202	Sequence 2202, Ap
32	22	1.1	10716	13	US-10-311-455-1392	Sequence 1392, App
33	22	1.1	13511	13	US-10-311-455-253	Sequence 253, App
34	22	1.1	14798	13	US-10-311-455-1005	Sequence 1005, Ap
35	22	1.1	40862	13	US-10-311-455-2045	Sequence 2045, Ap
36	22	1.1	73334	13	US-10-311-455-2097	Sequence 2097, Ap
37	22	1.1	3673778	13	US-10-312-841-2	Sequence 2, Appli
38	22	1.1	729	13	US-10-027-632-155448	Sequence 155448,
39	21	1.1	729	14	US-10-027-632-155448	Sequence 155448,
40	21	1.1	737	9	US-09-563-817-372	Sequence 372, App
41	21	1.1	771	15	US-10-172-086-66	Sequence 66, Appl
42	21	1.1	910	13	US-10-190-312A-86	Sequence 86, Appl
43	21	1.1	1201	13	US-10-027-632-124069	Sequence 124069,
44	21	1.1	1201	14	US-10-027-632-124069	Sequence 124069,
45	21	1.1	1201	14	US-10-027-632-124069	Sequence 124069,

ALIGNMENTS

RESULT 1

US-10-198-846-13016  
; Sequence 13016, Application US/10198846  
; Publication No. US2003009974A1  
; GENERAL INFORMATION:  
; APPLICANT: Lillie, James  
; APPLICANT: Xu, Yongyao  
; APPLICANT: Wang, Youzhen  
; APPLICANT: Steinmann, Kathleen  
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS  
; TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND  
; FILE REFERENCE: MRI-049  
; FILE REFERENCE: MRI-049  
; CURRENT APPLICATION NUMBER: US/10198.846  
; CURRENT FILING DATE: 2002-07-18  
; PRIOR APPLICATION NUMBER: 60/306,220  
; PRIOR FILING DATE: 2001-07-18  
; NUMBER OF SEQ ID NOS: 14084  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13016  
; LENGTH: 3470  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-198-846-13016

Query Match	91.6%	Score 1824;	DB 15;	Length 3470;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1824;	Conservative 0;	Mismatches 0;	Gaps 0;	
QY	147	GGACACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGATAGGTAGCAAAATTAAG	206	
Db	1647	GGACACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGATAGGTAGCAAAATTAAG	1706	
QY	207	GAGTATTCACAAAGTACCAACATGGAGGAGCTATGTTCCCTAAATTTATGGTGTAGCTGAAG	266	
Db	1707	GAGTATTCACAAAGTACCAACATGGAGGAGCTATGTTCCCTAAATTTATGGTGTAGCTGAAG	1766	
QY	267	GTGAAGATGATTTTATTGGAAGAGGTATTTGTTAATGGAAGATGATCTCCAGTGGAGG	326	



Db 1767 GTGAAGATGATTTATTGAGAGGATTTGTTGTAATGGAAGATGATTTCTCCAGTGGAG 1826  
QY 327 CTGAGAGCACCTTAATACTCCCGAAACCTTGTGCTGATGGAATAATTAGCAATTCATATG 386  
Db 1827 CTGAGAGCACCTTAATACTCCCGAAACCTTGTGCTGATGGAATAATTAGCAATTCATATG 1886  
QY 387 TAGACTTTTGGAGATCCCTCTCTGAAAGGAGGAGAGAAAGAAAGAAATTCCTGTGT 446  
Db 1887 TAGACTTTTGGAGATCCCTCTCTGAAAGGAGGAGAGAAAGAAAGAAATTCCTGTGT 1946  
QY 447 TTTGTAATGATGTTGAAAGAAATGATAGACAGCAGATTGGACAGCAGCCTGAACATTTGGT 506  
Db 1947 TTTGTAATGATGTTGAAAGAAATGATAGACAGCAGATTGGACAGCAGCCTGAACATTTGGT 2006  
QY 507 CTGCTATAGAGATATCTTGAATCTATGATCTATGATCTAATCAAACTAACCAATTTCAATG 566  
Db 2007 CTGCTATAGAGATATCTTGAATCTATGATCTAATCAAACTAACCAATTTCAATG 2066  
QY 567 GTGCAATTCCTGATGCCAGCTTCTTCTTAAGAGGATCAATTTGGCCCAAAATTTATGAAT 626  
Db 2067 GTGCAATTCCTGATGCCAGCTTCTTCTTAAGAGGATCAATTTGGCCCAAAATTTATGAAT 2126  
QY 627 TCTTAAAGTCAAGAGGAGAGGATTTCCAAAGATATCTACAGAACTTCTGAGCATCCAG 686  
Db 2127 TCTTAAAGTCAAGAGGAGAGGATTTCCAAAGATATCTACAGAACTTCTGAGCATCCAG 2186  
QY 745 RAATCAGTAATAGTCAACTTCTTGGCAGACTTCTTCCCTTAATTTGGGGGAAACACAAT 746  
Db 2187 RAATCAGTAATAGTCAACTTCTTGGCAGACTTCTTCCCTTAATTTGGGGGAAACACAAT 2246  
QY 747 TTTCTGTAAGATATCTACAGATTAATCTTGGGAAATTAATAATCTGTTCTCTGAA 806  
Db 2247 TTTCTGTAAGATATCTACAGATTAATCTTGGGAAATTAATAATCTGTTCTCTGAA 2306  
QY 807 RAATTAAGAGAGAGAGGATTTGGACCTTTTATCATCAATTTCAATTAATCTT 866  
Db 2307 RAATTAAGAGAGAGGATTTGGACCTTTTATCATCAATTTCAATTAATCTT 2366  
QY 925 GTGAGTCTCCAAAGCTTAAACCAAGTAGACAGAACTGACCAATTTCTCAGCCCTACTTTCAG 926  
Db 2367 GTGAGTCTCCAAAGCTTAAACCAAGTAGACAGAACTGACCAATTTCTCAGCCCTACTTTCAG 2426  
QY 927 AAAACAAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGAAATA 986  
Db 2427 AAAACAAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGAAATA 2486  
QY 987 CAGAGAGAGAGAGAGAGATTTATTTATGAGGATGATGACTGTAGAGGAGTCTATG 1046  
Db 2487 CAGAGAGAGAGAGAGAGATTTATTTATGAGGATGATGACTGTAGAGGAGTCTATG 2546  
QY 1047 ATTACTGATGATGATGAGAGGGTAGTTTCCAGGTTTCTGACTGGCTTCCATCTCT 1106  
Db 2547 ATTACTGATGATGATGAGAGGGTAGTTTCCAGGTTTCTGACTGGCTTCCATCTCT 2606  
QY 1107 TAATGGAACTCGAATCTCTTTAAAAACACCCCTGGAATGTATGATTAATCTTCT 1166  
Db 2607 TAATGGAACTCGAATCTCTTTAAAAACACCCCTGGAATGTATGATTAATCTTCT 2666  
QY 1167 AGTGTAACTAGAACAGCTATTCTTCAAGAGACCGTTTGGTCTCACTCAATCACTTCTCA 1226  
Db 2667 AGTGTAACTAGAACAGCTATTCTTCAAGAGACCGTTTGGTCTCACTCAATCACTTCTCA 2726  
QY 1227 GAGATGCTATTTCTGTGAAACACTGACCTGCTCTCTCCAGATGAGCAAAAGGAG 1286  
Db 2727 GAGATGCTATTTCTGTGAAACACTGACCTGCTCTCTCCAGATGAGCAAAAGGAG 2786  
QY 1287 CAAAACAGACTTTTGAAGAAATGATGATTAATCTCCAGATCTGTTAGTCAAGTGTATTG 1346  
Db 2787 CAAAACAGACTTTTGAAGAAATGATGATTAATCTCCAGATCTGTTAGTCAAGTGTATTG 2846  
QY 1347 GTGAAGAAACCAAGTATGAAGCATCAGATTTCTGTTTGGTCTTACAGCAACCAAGTAC 1406

Db 2847 GTGAAGAAACCAAGTATGAAGCATCAGATTTCTGTTGATGGCTTACAGCAACAGTAC 2906  
QY 1407 TCAACAAGCAGCTGACTTATGTTTATTGGACATTTGTGATACAGGAATCTGTTTCCAGAGC 1466  
Db 2907 TCAACAAGCAGCTGACTTATGTTTATTGGACATTTGTGATACAGGAATCTGTTTCCAGAGC 2966  
QY 1467 TCAATAAGGTACAAAGGAAGTTACTCTGTGACATCTTGGATGTAAACACTTGGATTTG 1526  
Db 2967 TCAATAAGGTACAAAGGAAGTTACTCTGTGACATCTTGGATGTAAACACTTGGATTTG 3026  
QY 1527 GTATAGATTAACCCATTTGAAATTTCTGCTGTGCGAGGGTGTAGAAATTTACTTTTTTGG 1586  
Db 3027 GTATAGATTAACCCATTTGAAATTTCTGCTGTGCGAGGGTGTAGAAATTTACTTTTTTGG 3086  
QY 1587 GTATATTTCTTATATATATATGATCGCTGTCTGAAATTTTACTTATTTTCTTTTTT 1646  
Db 3087 GTATATTTCTTATATATATATGATCGCTGTCTGAAATTTTACTTATTTTCTTTTTT 3146  
QY 1647 AATAAGACTAACACAAATTAATTAATAAGTGAATGAGTCTCATAGTCTTTTCATTTG 1706  
Db 3147 AATAAGACTAACACAAATTAATTAATAAGTGAATGAGTCTCATAGTCTTTTCATTTG 3206  
QY 1707 CTAGCTGTGATCCAAATTTTATAGAACATAAGTCACTTGTATTGCCATTTTAAAGA 1766  
Db 3207 CTAGCTGTGATCCAAATTTTATAGAACATAAGTCACTTGTATTGCCATTTTAAAGA 3266  
QY 1767 GAAATTTCTATATGATTTATGCGCAACAGATAGACTGTATAAATCTGTTGTATAGC 1826  
Db 3267 GAAATTTCTATATGATTTATGCGCAACAGATAGACTGTATAAATCTGTTGTATAGC 3326  
QY 1827 TTTGAAATTAATTAATGCTGTATGAGAAACAGGAATAAGATCTGATTTTCTTAGAGTT 1886  
Db 3327 TTTGAAATTAATTAATGCTGTATGAGAAACAGGAATAAGATCTGATTTTCTTAGAGTT 3386  
QY 1887 AATATATTTTAGTAGATGTTTCTTTTATTTTATTTTGTACATAGTTAACTGTGTATC 1946  
Db 3387 AATATATTTTAGTAGATGTTTCTTTTATTTTATTTTGTACATAGTTAACTGTGTATC 3446  
QY 1947 TATAATAAGCATCTCATATGAG 1970  
Db 3447 TATAATAAGCATCTCATATGAG 3470

## RESULT 2

US-10-117-722-922  
; Sequence 922, Application US/10117722  
; Publication No. US20030219744A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2BCIP  
; CURRENT APPLICATION NUMBER: US/10/117,722  
; CURRENT FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: 09/620,312  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1104  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (71)..(1543)

US-10-117-722-922

Query Match	76.9%	Score 1532;	DB 13;	Length 1716;
Best Local Similarity	100.0%;	Prod. No. 0;		
Matches 1532;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	147	GGAAACACACAGAAAAGGGAGAGATCATTTTGGAAATCAGACAGATAGGTAGCAGAAATTTAAAG	206	
Db	177	GGAAACACACAGAAAAGGGAGAGATCATTTTGGAAATCAGACAGATAGGTAGCAGAAATTTAAAG	236	
Qy	207	GAGTATTCAAAGGTACCACAATGGAGGAGCTATGTTGCCCTAAATATATGGTGTAGCTGAAG	266	
Db	237	GAGTATTCAAAGGTACCACAATGGAGGAGCTATGTTGCCCTAAATATATGGTGTAGCTGAAG	296	
Qy	267	GTGAGAGATGATTTTATTTGAAGAAGGTATTTGTTTAATGGAAGATGATTTCTCAGTGGAGG	326	
Db	297	GTGAGAGATGATTTTATTTGAAGAAGGTATTTGTTTAATGGAAGATGATTTCTCAGTGGAGG	356	
Qy	327	CTGTGAGCACACCTATATACCTCCCGAAACCTTCGTGCATCGAAAAATTTAGCATTTCCATATG	386	
Db	357	CTGTGAGCACACCTATATACCTCCCGAAACCTTCGTGCATCGAAAAATTTAGCATTTCCATATG	416	
Qy	387	TAGACITTTTTTGAGGATCCCTCTCTGAAAGGAGAGAAAAAGAAAGAAATTTCTGTGT	446	
Db	417	TAGACITTTTTTGAGGATCCCTCTCTGAAAGGAGAGAAAAAGAAAGAAATTTCTGTGT	476	
Qy	447	TTTGTATTGATTTGAAGAATAATATAGAGAGCAGTTGGACACGAGCCTGGAACATTTGGT	506	
Db	477	TTTGTATTGATTTGAAGAATAATATAGAGAGCAGTTGGACACGAGCCTGGAACATTTGGT	536	
Qy	507	CTGTCTATAGAGATATCTTTGMAATTTCTATGTTCTATGTAATCAAACTCAACAGATTTTCATG	566	
Db	537	CTGTCTATAGAGATATCTTTGMAATTTCTATGTTCTATGTAATCAAACTCAACAGATTTTCATG	596	
Qy	567	GTGCATTTTCTGATGCCAGCTTCCTTCTAAGAGGATCATTTGGCCCCCAAAAAATTTATGAAT	626	
Db	597	GTGCATTTTCTGATGCCAGCTTCCTTCTAAGAGGATCATTTGGCCCCCAAAAAATTTATGAAT	656	
Qy	627	TCCTTAAAGTCAAAGAGGAGAGTTCCAAGATAATCTACAGAAATTTCTCGACAGATCCAG	686	
Db	657	TCCTTAAAGTCAAAGAGGAGAGTTCCAAGATAATCTACAGAAATTTCTCGACAGATCCAG	716	
Qy	687	AACGTAGTAAATAGTCAACTTCTGGCAGACTTTCTTTCCCTTAATATGGTGGGGAACACAAAT	746	
Db	717	AACGTAGTAAATAGTCAACTTCTGGCAGACTTTCTTTCCCTTAATATGGTGGGGAACACAAAT	776	
Qy	747	TTCTTGATATAGATACTACCAGATGTAATCTTTGGGAAAAATATATAAAATCTGTTCTCGAA	806	
Db	777	TTCTTGATATAGATACTACCAGATGTAATCTTTGGGAAAAATATATAAAATCTGTTCTCGAA	836	
Qy	807	AACTAATGAAGAGAGAAAGGTACAGCATTTGGAAACCTTTTATCATGAAATTTCAATTAATCTTT	866	
Db	837	AACTAATGAAGAGAGAAAGGTACAGCATTTGGAAACCTTTTATCATGAAATTTCAATTAATCTTT	896	
Qy	867	GTGAGTCTCCAAAGCCTTAAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACTTTGAG	926	
Db	897	GTGAGTCTCCAAAGCCTTAAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACTTTGAG	956	
Qy	927	AAAAACAACAGAGCTTTTCAATGATCTGTTTAAAAATATATGCAAAACCGTGTCTGAAAAATA	986	
Db	957	AAAAACAACAGAGCTTTTCAATGATCTGTTTAAAAATATATGCAAAACCGTGTCTGAAAAATA	1016	
Qy	987	CAGAGAGAAACGAAAAATCAGAAATTTTATATGAGGAGTATGACATGTATAGAGGAGTCTATG	1046	
Db	1017	CAGAGAGAAACGAAAAATCAGAAATTTTATATGAGGAGTATGACATGTATAGAGGAGTCTATG	1076	
Qy	1047	ATTACCTGATGATATGTAGAGCGGGTAGTTTTTCCAGGTTTCTGTACTGGGTTTCATCATCTCT	1106	
Db	1077	ATTACCTGATGATATGTAGAGCGGGTAGTTTTTCCAGGTTTCTGTACTGGGTTTCATCATCTCT	1136	
Qy	1107	TAATGGGAACTCGAAATTCCTTTTAAAAACAACCCCTGGAAATGTATATCTGATTTACTATCTTC	1166	
Db	1137	TAATGGGAACTCGAAATTCCTTTTAAAAACAACCCCTGGAAATGTATATCTGATTTACTATCTTC	1196	

RESULT 3

```

RES001 3
US-10-037-270-922
; Sequence 922, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyun
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunging
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1e1 Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: Pt_Fl_genes version 1.0
; SEQ ID NO 922
; LENGTH: 1716

```

1077	ATTACCTGATGATATGTAGGACGGGTAGTTTCCAGGTTCTCTGACTGGCTTCATCATCTCT	1130
1107	TAATGGGAACCTCGAATCCTCTTTTAAACACACCTCGAAATGTATGATTAATACTATCTTC	1166
1137	TAATGGGAACCTCGAATCCTCTTTTAAACACACCTCGAAATGTATGATTAATACTATCTTC	1196
1167	AGTGTAACTAGACACACTATTTTCAGGAGCAGCGTTTGGTCTCCTCACTCATAAACACTCTCA	1226
1197	AGTGTAACTAGACACACTATTTTCAGGAGCAGCGTTTGGTCTCCTCACTCATAAACACTCTCA	1256
1227	GAGATGCTATATCTGTGAAACACACTCGCTCTCTCCAAAGATAGCAAAAGGAG	1286
1257	GAGATGCTATATCTGTGAAACACACTCGCTCTCTCCAAAGATAGCAAAAGGAG	1316
1287	CAAAACAGACACTTTTGAAGAAATGATGAATTTACATTCAGATCTGTTAGTCAAGTGTTTG	1346
1317	CAAAACAGACACTTTTGAAGAAATGATGAATTTACATTCAGATCTGTTAGTCAAGTGTTTG	1376
1347	GTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTTACAGCAACCAAGTAC	1406
1377	GTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTTACAGCAACCAAGTAC	1436
1407	TCAACCAAGCAGCTGACCTATGTTTATTGACATTTGTGATACAGGAACCTGTTTCCAGAGC	1466
1437	TCACCAAGCAGCTGACCTATGTTTATTGACATTTGTGATACAGGAACCTGTTTCCAGAGC	1496
1467	TCAATAAGGTTCAAAAGGAAGTTTACCTCTCTGATCATCTTGGATGTAAACACTTTGTTTG	1526
1497	TCAATAAGGTTCAAAAGGAAGTTTACCTCTCTGATCATCTTGGATGTAAACACTTTGTTTG	1556
1527	GTATAGATTAACCAATTTGAAATTTCTGCTGTGGAGGGTGTAGAAATTTACTTTTGTG	1586
1557	GTATAGATTAACCAATTTGAAATTTCTGCTGTGGAGGGTGTAGAAATTTACTTTTGTG	1616
1587	GTATATCTTATATATATATGATCATGCTGTCTGAAATTTTATTTTGTGTTT	1646
1617	GTATATCTTATATATATATGATCATGCTGTCTGAAATTTTATTTTGTGTTT	1676
1647	ATAAAGACTAACCAACTTAAATGATTAATAA	1678
1677	ATAAAGACTAACCAACTTAAATGATTAATAA	1708
RESULT 4		
US-10-117-922-923		
; Sequence 923, Application US/10117722		
; Publication No. US20030219744A1		
; GENERAL INFORMATION:		
; APPLICANT: Tang, Y. Tom		
; APPLICANT: Liu, Chenghua		
; APPLICANT: Asundi, Vinod		
; APPLICANT: Zhang, Jie		
; APPLICANT: Drmanac, Radoje T.		
; TITLE OF INVENTION: No. US20030219744A1elel Nucleic Acids and		
; TITLE OF INVENTION: Polypeptides		
; FILE REFERENCE: 784CIP2BCIP		
; CURRENT APPLICATION NUMBER: US/10/117,722		
; CURRENT FILING DATE: 2002-04-04		
; PRIOR APPLICATION NUMBER: 09/620,312		
; PRIOR FILING DATE: 2000-07-19		
; PRIOR APPLICATION NUMBER: 09/552,317		
; PRIOR FILING DATE: 2000-04-25		
; PRIOR APPLICATION NUMBER: 09/488,725		
; PRIOR FILING DATE: 2000-01-21		
; NUMBER OF SEQ ID NOS: 1104		
; SOFTWARE: pt_FL_genes Version 1.0		
; SEQ ID NO 923		
; LENGTH: 1551		
; TYPE: DNA		
; ORGANISM: Homo sapiens		
; FEATURE:		
; NAME/KEY: CDS		
; LOCATION: (71)..(1378)		

RESULT 4  
US-10-117-722-923  
; Sequence 923, Application US/10117722  
; Publication No. US20030219744A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenchua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. US20030219744A1el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2BCIP  
; CURRENT APPLICATION NUMBER: US/10/117,722  
; CURRENT FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: 09/620,312  
; PRIOR FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1104  
; SOFTWARE: pc FL\_genes Version 1.0  
; SEQ ID NO 923  
; LENGTH: 1551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (71)..(1378)

US-10-117-722-923

```
Query Match 46.1%; Score 918; DB 13; Length 1551;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 GGACACACAGAAAGGGAGAGATCAATTTGGAATCAGCAGAAATAGGTAGCAAAATTTAAAG 206
DB 177 GGACACACAGAAAGGGAGAGATCAATTTGGAATCAGCAGAAATAGGTAGCAAAATTTAAAG 236
QY 207 GAGTATTCAAAGTACCAATGAGGAGAGTATGTTGTAATGGAAGATGATCTCCAGTGGAG 266
DB 237 GAGTATTCAAAGTACCAATGAGGAGAGTATGTTGTAATGGAAGATGATCTCCAGTGGAG 296
QY 267 GTGAGATGATTTTATTGAAGAGGATGTTGTTGTAATGGAAGATGATCTCCAGTGGAG 326
DB 297 GTGAGATGATTTTATTGAAGAGGATGTTGTTGTAATGGAAGATGATCTCCAGTGGAG 356
QY 327 CTGTGAGCACACCTTAATCTCCCGAAACCTTGCTGATGGAATTTAGCATTTCCATATG 386
DB 357 CTGTGAGCACACCTTAATCTCCCGAAACCTTGCTGATGGAATTTAGCATTTCCATATG 416
QY 387 TAGACTTTTTGAGGATCCCTCTCTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 446
DB 417 TAGACTTTTTGAGGATCCCTCTCTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 476
QY 447 TTTGTTATGATTTGAAAGAAATGATAGAGAGCAGTGGACAGAGCCTTGAACATTTGTT 506
DB 477 TTTGTTATGATTTGAAAGAAATGATAGAGAGCAGTGGACAGAGCCTTGAACATTTGTT 536
QY 507 CTGTCTATAGAGATCTTGAATCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 566
DB 537 CTGTCTATAGAGATCTTGAATCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 596
QY 567 GTGCATTTCTGATGCCAGCTTCCTTCTAAGAGGATCATTTGGCCCGCAAAATTTATGAAT 626
DB 597 GTGCATTTCTGATGCCAGCTTCCTTCTAAGAGGATCATTTGGCCCGCAAAATTTATGAAT 656
QY 627 TCTTAAAGTCAAAGAGGAGAGGATTTCAAGAAATATCTACAGAACTTTCTGAGATCCAG 686
DB 657 TCTTAAAGTCAAAGAGGAGAGGATTTCAAGAAATATCTACAGAACTTTCTGAGATCCAG 716
QY 687 AACTGATTAATGATCACTTCTGAGAGCTTTCTTCCCTTAATGTTGGGGAACACCAAT 746
DB 717 AACTGATTAATGATCACTTCTGAGAGCTTTCTTCCCTTAATGTTGGGGAACACCAAT 776
QY 747 TCTTGAATAGATATCACTACAGATCTAAATCTTGGGAAATTTATAAATCTGTTCCCTGGA 806
DB 777 TCTTGAATAGATATCACTACAGATCTAAATCTTGGGAAATTTATAAATCTGTTCCCTGGA 836
QY 807 AACTTAATGAAGAGAGGATCAGATTTGGAACCTTTTATCATGAATTTTCAATTAATCTTT 866
DB 837 AACTTAATGAAGAGAGGATCAGATTTGGAACCTTTTATCATGAATTTTCAATTAATCTTT 896
QY 867 GTGAGTCTCAAAGCCTTAACCAAGTACACAGAACTGACCATTTCTCAGCCCTTACTTCAG 926
DB 897 GTGAGTCTCAAAGCCTTAACCAAGTACACAGAACTGACCATTTCTCAGCCCTTACTTCAG 956
QY 927 AARACACAGAGCCTTTCAATGATCTGTTTAAATTAATGCAACCGTCTGGAATA 986
DB 957 AARACACAGAGCCTTTCAATGATCTGTTTAAATTAATGCAACCGTCTGGAATA 1016
QY 987 CAGAGAGAGCAAAATCAGAAATTTTATGAGGATGATGATCTGTAGAGGAGTCTATG 1046
DB 1017 CAGAGAGAGCAAAATCAGAAATTTTATGAGGATGATGATCTGTAGAGGAGTCTATG 1076
QY 1047 ATTACCTGATGATGTAG 1064
DB 1077 ATTACCTGATGATGTAG 1094
```

RESULT 5

US-10-037-270-923

```
; Sequence 923, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Dmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 923
; LENGTH: 1551
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71)..(1378)
; US-10-037-270-923
```

```
Query Match 46.1%; Score 918; DB 15; Length 1551;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 GGACACACAGAAAGGGAGAGATCAATTTGGAATCAGCAGAAATAGGTAGCAAAATTTAAAG 206
DB 177 GGACACACAGAAAGGGAGAGATCAATTTGGAATCAGCAGAAATAGGTAGCAAAATTTAAAG 236
QY 207 GAGTATTCAAAGTACCAATGAGGAGAGTATGTTGTAATGGAAGATGATCTCCAGTGGAG 266
DB 237 GAGTATTCAAAGTACCAATGAGGAGAGTATGTTGTAATGGAAGATGATCTCCAGTGGAG 296
QY 267 GTGAGATGATTTTATTGAAGAGGATGTTGTTGTAATGGAAGATGATCTCCAGTGGAG 326
DB 297 GTGAGATGATTTTATTGAAGAGGATGTTGTTGTAATGGAAGATGATCTCCAGTGGAG 356
QY 327 CTGTGAGCACACCTTAATCTCCCGAAACCTTGCTGATGGAATTTAGCATTTCCATATG 386
DB 357 CTGTGAGCACACCTTAATCTCCCGAAACCTTGCTGATGGAATTTAGCATTTCCATATG 416
QY 387 TAGACTTTTTGAGGATCCCTCTCTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 446
DB 417 TAGACTTTTTGAGGATCCCTCTCTGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 476
QY 447 TTTGTTATGATTTGAAAGAAATGATAGAGAGCAGTGGACAGAGCCTTGAACATTTGTT 506
DB 477 TTTGTTATGATTTGAAAGAAATGATAGAGAGCAGTGGACAGAGCCTTGAACATTTGTT 536
QY 507 CTGTCTATAGAGATCTTGAATCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 566
DB 537 CTGTCTATAGAGATCTTGAATCTCTGATCTGATCTGATCTGATCTGATCTGATCTGATCT 596
QY 567 GTGCATTTCTGATGCCAGCTTCCTTCTAAGAGGATCATTTGGCCCGCAAAATTTATGAAT 626
```

597 GTGCAATTCCTGATGCCAGCTTCTCTAAGAGATCAATGGCCCCAATAATATGAT 656  
627 TCTTAAAGTCAAAGAGGAAGTTCAGAAATATCTACAGAACTTCGAGCATCCAG 686  
657 TCTTAAAGTCAAAGAGGAAGTTCAGAAATATCTACAGAACTTCGAGCATCCAG 716  
687 AACTGAGTAATAGTCACTCTCGAGACTTCTTCCCTTAAGTGGGGGAACAAAT 746  
717 AACTGAGTAATAGTCACTCTCGAGACTTCTTCCCTTAAGTGGGGGAACAAAT 776  
747 TCTTGAAGATCTACAGATGTAATCTTGGGAAAATATATAAAATCTGTCCTGGAA 806  
777 TCTTGAAGATCTACAGATGTAATCTTGGGAAAATATATAAAATCTGTCCTGGAA 836  
807 AACTAATGAAGAGAAAGTTCAGATTTGGAACCTTTTATCATGAATTCATTAATCTT 866  
837 AACTAATGAAGAGAAAGTTCAGATTTGGAACCTTTTATCATGAATTCATTAATCTT 896  
867 GTGAGTCTCAAAGCTTAAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTCAG 926  
897 GTGAGTCTCAAAGCTTAAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTCAG 956  
927 AAACAACAAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAAAATA 986  
957 AAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAAAATA 1016  
987 CAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTAGAAGGATCTATG 1046  
1017 CAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTAGAAGGATCTATG 1076  
1047 ATTACTGATGATGATG 1064  
1077 ATTACTGATGATGATG 1094

## RESULT 6

US-09-918-995-22530  
; Sequence 22530, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 22530  
; LENGTH: 535  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)..(535)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-22530

Query Match 22.3%; Score 444; DB 11; Length 535;  
Best Local Similarity 100.0%; Pred. No. 5.5e-206;  
Matches 444; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1393 ACAGCAACAGTACTCAACAGAGAGTCACTTATGTTTATGACATTTGATACAGGA 1452  
34 ACAGCAACAGTACTCAACAGAGAGTCACTTATGTTTATGACATTTGATACAGGA 93  
1453 ACTGTTTCCAGGCTCAATAGGTACAAAGAGTACCTCTGACATCTGGATGA 1512  
94 ACTGTTTCCAGGCTCAATAGGTACAAAGAGTACCTCTGACATCTGGATGA 153

1513 AACACTTGGATTGGTATAGATAACCCATTGAAATTTCTGCTGTGCGAGGGTGTAGAA 1572  
154 AACACTTGGATTGGTATAGATAACCCATTGAAATTTCTGCTGTGCGAGGGTGTAGAA 213  
1573 ATTACTTTTTGGGTATATTTCTTATATATATATATATATATATATATATATATAT 1632  
214 ATTACTTTTTGGGTATATTTCTTATATATATATATATATATATATATATATATAT 273  
1633 TATTTTTTTGTTTTTAAATAAGACTAACCAAACTTAAATGATTAAGAAGTATGAGTCTCA 1692  
274 TATTTTTTTGTTTTTAAATAAGACTAACCAAACTTAAATGATTAAGAAGTATGAGTCTCA 333  
1693 TAGTCTTTCAATTTGCTAGCTGTGATCCAAATTTTATTAAGAACATAAGTCACTTGTATTG 1752  
334 TAGTCTTTCAATTTGCTAGCTGTGATCCAAATTTTATTAAGAACATAAGTCACTTGTATTG 393  
1753 CCATTTTTAAAGAGAAAATTCATATGATGTTATGCGAAAACAGATAAGACTGATAAAT 1812  
394 CCATTTTTAAAGAGAAAATTCATATGATGTTATGCGAAAACAGATAAGACTGATAAAT 453  
1813 TCGTATTGATAGCTTTGAAAATA 1836  
454 TCGTATTGATAGCTTTGAAAATA 477

RESULT 7  
US-09-918-995-20205  
; Sequence 20205, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 20205  
; LENGTH: 508  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-20205

Query Match 21.0%; Score 419; DB 11; Length 508;  
Best Local Similarity 100.0%; Pred. No. 8.9e-194;  
Matches 419; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
1190 CAGGAGCACCGTTTGGTCTCACTATAACACTTCTCAGAGATGCTATATTTCTGTGAAAAC 1249  
65 CAGGAGCACCGTTTGGTCTCACTATAACACTTCTCAGAGATGCTATATTTCTGTGAAAAC 124  
1250 ACTGAACCTCGCTCTCTCAGAGATAAGCAAAAAGAGCAAAAACAGACTTTTGAAGAAATG 1309  
125 ACTGAACCTCGCTCTCTCAGAGATAAGCAAAAAGAGCAAAAACAGACTTTTGAAGAAATG 184  
1310 ATGAATTACATTCAGATCTGTTAGTCAAGTGTATTTGTTGAAGAAACCAAGTATCAAGC 1369  
185 ATGAATTACATTCAGATCTGTTAGTCAAGTGTATTTGTTGAAGAAACCAAGTATCAAGC 244  
1370 ATCAGACTTCTGTTTGTAGTGGCTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTT 1429  
245 ATCAGACTTCTGTTTGTAGTGGCTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTT 304  
1430 TTATTGGACATTTGATACAGAACTGTTTCAGAGCTCAATAAGGTACAAAGGAAGTT 1489  
305 TTATTGGACATTTGATACAGAACTGTTTCAGAGCTCAATAAGGTACAAAGGAAGTT 364  
1490 ACCTCTGTGACATCTTGGATGTAACACTTGGATTTGTTAGTATAGATTAACCCATTCGAAT 1549  
365 ACCTCTGTGACATCTTGGATGTAACACTTGGATTTGTTAGTATAGATTAACCCATTCGAAT 424

QY 1550 TCTGCTGTCGAGGGTGGTAGAAATTTACTTTTTGGGATATTTCTTATATATATATG 1608  
Db 425 TCTGCTGTCGAGGGTGGTAGAAATTTACTTTTTGGGATATTTCTTATATATATATG 483

RESULT 8

US-09-918-995-5183  
; Sequence 5183, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; CURRENT FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5183  
; LENGTH: 397  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-5183

Query Match 19.9%; Score 397; DB 11; Length 397;  
Best Local Similarity 100.0%; Pred. No. 4.8e-183;  
Matches 397; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1044 ATGATTACCTGATGTATGACGCGGTAGTTTCCAGGTTCTGCTGCTGCTTCAATC 1103  
Db 1 ATGATTACCTGATGTATGACGCGGTAGTTTCCAGGTTCTGCTGCTGCTTCAATC 60  
QY 1104 TCTTAATGGAACTCGAATCTCTTTTAAACACACCTGGAATGATATGATTAATC 1163  
Db 61 TCTTAATGGAACTCGAATCTCTTTTAAACACACCTGGAATGATATGATTAATC 120  
QY 1164 TCCAGTGAACCTAGAACAGCTATTTCCAGGAGCAGCGTTGGTCTCACTATAACCTTC 1223  
Db 121 TCCAGTGAACCTAGAACAGCTATTTCCAGGAGCAGCGTTGGTCTCACTATAACCTTC 180  
QY 1224 TCAGAGATGCTATTTCTGTAACACACCTGCTGCTCTCCAGATAGCAAAAG 1283  
Db 181 TCAGAGATGCTATTTCTGTAACACACCTGCTGCTCTCCAGATAGCAAAAG 240  
QY 1284 GAGCAAAACAGACTTTTGAAGAAATGATGAATTCAGATCTGTTAGTCAAGTGTA 1343  
Db 241 GAGCAAAACAGACTTTTGAAGAAATGATGAATTCAGATCTGTTAGTCAAGTGTA 300  
QY 1344 TTGGTGAAGAAACCAAGTATGAAGCATCAGACTCTGTTGATGCTTTACAGCAACAG 1403  
Db 301 TTGGTGAAGAAACCAAGTATGAAGCATCAGACTCTGTTGATGCTTTACAGCAACAG 360  
QY 1404 TACTCAACAGGAGCTGACTATGTTTATGACAT 1440  
Db 361 TACTCAACAGGAGCTGACTATGTTTATGACAT 397

RESULT 9

US-09-954-531-826/c  
; Sequence 826, Application US/09954531  
; Patent No. US20020165180A1  
; GENERAL INFORMATION:  
; APPLICANT: Weaver, Zoe  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand  
; FILE REFERENCE: 689290-77  
; CURRENT APPLICATION NUMBER: US/09/954,531  
; CURRENT FILING DATE: 2002-05-02  
; PRIOR APPLICATION NUMBER: US/60/233,133  
; PRIOR FILING DATE: 2000-09-18

; PRIOR APPLICATION NUMBER: US/60/234,009  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,034  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,509  
; PRIOR FILING DATE: 2000-09-22  
; PRIOR APPLICATION NUMBER: US/60/234,567  
; PRIOR FILING DATE: 2000-09-22  
; NUMBER OF SEQ ID NOS: 1392  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 826  
; LENGTH: 451  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (1)...(451)  
; OTHER INFORMATION: n=a,t,g or c  
US-09-954-531-826

Query Match 19.3%; Score 385; DB 10; Length 451;  
Best Local Similarity 100.0%; Pred. No. 3.6e-177;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1597 ATATATATTATGTACATCGCTGCTGAAATTTTGTAGTTATTTTGTATTAATAAGACT 1656  
Db 385 ATATATATTATGTACATCGCTGCTGAAATTTTGTAGTTATTTTGTATTAATAAGACT 326  
QY 1657 AACCAAACTTAATGATTAAGTGAATGATGCTCATAGTCTTCAATTTGCTAGCTGGA 1716  
Db 325 AACCAAACTTAATGATTAAGTGAATGATGCTCATAGTCTTCAATTTGCTAGCTGGA 266  
QY 1717 TCCAAATTTTATAGACATAAGTCACTGTTGTTATTCCTATTTTAAAGAGAAATTCAT 1776  
Db 265 TCCAAATTTTATAGACATAAGTCACTGTTGTTATTCCTATTTTAAAGAGAAATTCAT 206  
QY 1777 AATGATCTTATGCAACACAGATAAGACTGATAAACTTCGTTATGTATAGCTTTGAAATA 1836  
Db 205 AATGATCTTATGCAACACAGATAAGACTGATAAACTTCGTTATGTATAGCTTTGAAATA 146  
QY 1837 ATTATGCTAGTAGGAGAAACAGGAATAAGATCTGATTTCTTAGAGTTAATATTTT 1896  
Db 145 ATTATGCTAGTAGGAGAAACAGGAATAAGATCTGATTTCTTAGAGTTAATATATTTT 86  
QY 1897 AGTAGATGTTTTTCCCTTTTTTTTTTTTATTTTGTACATAGTTAACTGTGTATCTATAATAA 1956  
Db 85 AGTAGATGTTTTTCCCTTTTTTTTTTTTATTTTGTACATAGTTAACTGTGTATCTATAATAA 26  
QY 1957 GCATCCTATATAGTGTTTTAAATAAT 1981  
Db 25 GCATCCTATATAGTGTTTTAAATAAT 1

RESULT 10

US-10-242-535A-25415  
; Sequence 25415, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liow, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005  
; CURRENT APPLICATION NUMBER: US/10/242,535A  
; CURRENT FILING DATE: 2002-09-12  
; PRIOR APPLICATION NUMBER: US 10/085,783  
; PRIOR FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994



	Query Match	17.2%;	Score 342;	DB 12;	Length 446;
	Best Local Similarity	100.0%;	Pred. No. 3.6e-156;		
	Matches 342;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1226	AGAGATGCTATATTCTGTGAAACACTGAACTCGCTCTCTCCAAGATAAGCAAAAAGGA	1285		
Db	1	AGAGATGCTATATTCTGTGAAACACTGAACTCGCTCTCTCCAAGATAAGCAAAAAGGA	60		
QY	1286	GCAAAACAGACATTTTGAGAAATGATGAATTCATTCAGATCTGTTTACTCAAGTGTAAT	1345		
Db	61	GCAAAACAGACATTTTGAGAAATGATGAATTCATTCAGATCTGTTTACTCAAGTGTAAT	120		

QY	1346	QSIAGAAACCCAGATGAAAGCAACAACCTCTGTGTTTGGATGCGAACCAGTA	1490
Db	121	GGTGAAGAACCAAGTATGAAGCATCAGACTCTGTTTGATGGCTTACAGCAACCAATA	180
QY	1406	CTCACAGCAGCTGACTTATTGTTTTATTGGACATGTGATACAGGAACCTGTTCCAGAG	1465
Db	181	CTCAACAGCAGCTGACTTATTGTTTTATTGGACATGTGATACAGGAACCTGTTCCAGAG	240
QY	1466	CTCAATAAGGTACAAAAGGAAGTTACCTCTGTGCACATCTTGAATGTAAACACTTGGAATT	1525
Db	241	CTCAATAAGGTACAAAAGGAAGTTACCTCTGTGCACATCTTGAATGTAAACACTTGGAATT	300
QY	1526	GGTATGAATAACCCATTGAAATTTTCGTGTGCGAGGGTGG	1567
Db	301	GGTATGAATAACCCATTGAAATTTTCGTGTGCGAGGGTGG	342

```

RESULT 12
US-09-918-995-17035
; Sequence 17035, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17035
; LENGTH: 464
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(464)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-17035

```

Query Match 4.9%; Score 98; DB 11; Length 464;  
Best Local Similarity 100.0%; Pred. No. 5.4e-37;  
Matches 98; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1887 AATATATTTTAGAGAGATTGGTTTTCCTTTTTTTATTTTGTACATGATTAACTGTGTATC 1946

Db 187 AATATATTTAGTAGATTGGTTTCTCTTTTATTTTGTACATAGTAACTGTGTATC 246  
Qy 1947 TATAATAAGCATCTCATATGAGTTTAAATAATAA 1984  
Db 247 TATAATAAGCATCTCATATGAGTTTAAATAATAA 284

## RESULT 13

US-09-908-975-12718  
; Sequence 12718, Application US/09908975  
; Publication No. US20030165843A1  
; GENERAL INFORMATION:  
; APPLICANT: SHOSHAN, AVI  
; APPLICANT: WASSERMAN, ALON  
; APPLICANT: MINTZ, ELI  
; APPLICANT: MINTZ, LIAT  
; APPLICANT: FAIGLER, SIMCHON  
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLICE  
; FILE REFERENCE: 36688-0005  
; CURRENT APPLICATION NUMBER: US/09/908,975  
; PRIOR FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: US 60/287,724  
; PRIOR FILING DATE: 2001-05-02  
; PRIOR APPLICATION NUMBER: US 60/221,607  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 32337  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 12718  
; LENGTH: 60  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-908-975-12718

Query Match 3.0%; Score 60; DB 13; Length 60;  
Best Local Similarity 100.0%; Pred. No. 1.6e-18;  
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 891 GTAGACAGACTGACCATCTCAGCCCTACTTCAGAAACACACAGAGCTTTTCAATG 950  
Db 1 GTAGACAGACTGACCATCTCAGCCCTACTTCAGAAACACACAGAGCTTTTCAATG 60

## RESULT 14

US-09-983-965-355  
; Sequence 355, Application US/09983965  
; Patent No. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: WARREN, WESLEY C.  
; APPLICANT: TAO, NENGBOING  
; APPLICANT: BYATT, JOHN C.  
; APPLICANT: MATHIALAGAN, NAGAPPAN  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; PRIOR FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 355  
; LENGTH: 305  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 42-BOVMS1-019-Q1-E1-C6  
US-09-983-965-355

Query Match 3.0%; Score 59; DB 10; Length 305;  
Best Local Similarity 100.0%; Pred. No. 5.8e-18;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 770 GTAAATCTTGGGAAAATTATAAAATCTGTTCTCTGGAAAACATAATGAAAGAAAGGTCA 828  
Db 118 GTAAATCTTGGGAAAATTATAAAATCTGTTCTCTGGAAAACATAATGAAAGAAAGGTCA 176

## RESULT 15

US-09-983-965-540  
; Sequence 540, Application US/09983965  
; Patent No. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: WARREN, WESLEY C.  
; APPLICANT: TAO, NENGBOING  
; APPLICANT: BYATT, JOHN C.  
; APPLICANT: MATHIALAGAN, NAGAPPAN  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; PRIOR FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 540  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (339)  
; OTHER INFORMATION:  
; OTHER INFORMATION: Clone ID: 64-BOVMS1-017-Q1-E1-H12  
US-09-983-965-540

Query Match 3.0%; Score 59; DB 10; Length 402;  
Best Local Similarity 100.0%; Pred. No. 5.9e-18;  
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 770 GTAAATCTTGGGAAAATTATAAAATCTGTTCTCTGGAAAACATAATGAAAGAAAGGTCA 828  
Db 45 GTAAATCTTGGGAAAATTATAAAATCTGTTCTCTGGAAAACATAATGAAAGAAAGGTCA 103

Search completed: January 31, 2004, 21:44:01  
Job time : 716 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 05:18:46 : Search time 5257 Seconds  
(without alignments)  
3618.599 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 465

Sequence: 1 MYLHFCLFRTYKRGESF.....ELFPELNKVKQEVTSVSM 465

Scoring table: OLIGO

Xgapop 60.0, Xgapext 60.0

Ygapop 60.0, Ygapext 60.0

Zgapop 6.0, Zgapext 7.0

Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5773148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh

-O=/cgn2 1/USPTO.spool/US09744313/runat 30012004 113612 4590/app query.fasta\_1.647

-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0 -LOOPEXT=0

-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45

-DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc

-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000

-USBR=US09744313 @CGN 1 1 3508 @runat 30012004 113612 4590 -NCPU=6 -ICPU=3

-NO\_WMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG

-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6

-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

GenEmbl:  
1: gb\_ba:\*  
2: gb\_hgt:\*  
3: gb\_in:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pi:\*  
9: gb\_pr:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vi:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_nu:\*  
20: em\_on:\*  
21: em\_or:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*

29: em\_vi:\*  
30: em\_hgt\_hum:\*  
31: em\_hgt\_inv:\*  
32: em\_hgt\_other:\*  
33: em\_hgt\_mus:\*  
34: em\_hgt\_pln:\*  
35: em\_hgt\_rod:\*  
36: em\_hgt\_mam:\*  
37: em\_hgt\_vrt:\*  
38: em\_sy:\*  
39: em\_hgtgo\_hum:\*  
40: em\_hgtgo\_mus:\*  
41: em\_hgtgo\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	465	100.0	1992	6	AX054819 Sequence
2	455	97.8	2661	9	AY044865 Homo sapi
3	455	97.8	3038	9	BC005110 Homo sapi
4	390	83.9	2925	9	BC048520 Homo sapi
5	374	80.4	1593	9	AF121863 Homo sapi
6	368	79.1	3576	9	AK026479 Homo sapi
7	330	71.0	3145	6	AX512835 Sequence
8	330	71.0	3145	9	AK000362 Homo sapi
9	269	57.8	3616	9	AK095380 Homo sapi
10	163	35.1	968	9	HS4420581 Homo sapi
11	141	30.3	425	6	BD113855 EST and e
12	119	25.6	358	6	BD026500 Sequence
13	109	23.4	1782	10	BC043328 Mus muscu
14	67	14.4	123010	9	AL589666 Human DNA
15	67	14.4	278898	2	AC118713 Mus muscu
16	59	12.7	224551	2	AC111832 Rattus no
17	55	11.8	278375	2	AC130093 Rattus no
18	47	10.1	454	11	G30543 human STS S
19	25	5.4	311	6	AX185131 Sequence
20	25	5.4	313	6	AX186466 Sequence
21	24	5.2	299	6	AX188027 Sequence
22	21	4.5	183972	2	BX537259 Danio rer
23	12	2.6	36154	2	AC100446 Mus muscu
24	11	2.4	160411	2	AC135935 Rattus no
25	11	2.4	200417	2	AC134192 Rattus no
26	11	2.4	230404	2	AC097160 Rattus no
27	10	2.2	214222	2	AC074160 Mus muscu
28	9	1.9	188	6	BD123051 EST and e
29	9	1.9	347	6	BD123053 EST and e
30	9	1.9	367	6	BD123050 EST and e
31	9	1.9	377	6	BD123049 EST and e
32	9	1.9	506	6	BD123048 EST and e
33	9	1.9	543	6	BD123052 EST and e
34	9	1.9	684	8	AF513658 Aphanoal
35	9	1.9	815	5	AF390220 Sarda sar
36	9	1.9	821	5	AF390239 Sarda sar
37	9	1.9	824	5	AF390246 Sarda sar
38	9	1.9	830	5	AF390225 Sarda sar
39	9	1.9	839	5	AF390245 Sarda sar
40	9	1.9	845	5	AF390232 Sarda sar
41	9	1.9	846	5	AF390235 Sarda sar
42	9	1.9	847	5	AF390212 Sarda sar
43	9	1.9	847	5	AF390216 Sarda sar
44	9	1.9	847	5	AF390217 Sarda sar
45	9	1.9	847	5	AF390217 Sarda sar

# ALIGNMENTS

RESULT 1

```

AX054819
LOCUS AX054819 1992 bp DNA linear PAT 13-JAN-2001
DEFINITION Sequence 3 from Patent WO0073334.
ACCESSION AX054819
VERSION AX054819.1 GI:12228268
KEYWORDS Homo sapiens (human)
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 Yue, H., Tang, Y. T. and Azimzai, Y.
AUTHORS Human sorting nexins
TITLE Patent: WO 0073334-A 3 07-DEC-2000;
JOURNAL Incyte Genomics, Inc. (US)
FEATURES
Location/Qualifiers
1..1992
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 671 a 308 c 372 g 641 t
ORIGIN

Alignment Scores:
Pred. No.: 0 Length: 1992
Score: 465.00 Matches: 465
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-09-744-313A-1 (1-465) x AX054819 (1-1992)
QY 1 MetTyrLeuIleHisPheCysLeuIlePheArgAsnThrGlnLysArgGlyValSerPhe 20
DB 116 AUGTACTTGATACATTTTGGTTTATATATTCAGACACACACAGAAAGGGAGATCATTT 175
QY 21 GlyIleSerArgIleGlySerIleLysGlyValPheLysSerThrMetGluGly 40
DB 176 GGAATCAGCAGATAGTAGCAAAATTAAGGAGTATTCAAAAGTACCACCAATGGAGGA 235
QY 41 AlaMetLeuProAsnTyrGlyValAlaGluGlyGluAspPheIleGluGlyIle 60
DB 236 GCTATCTGCCTAAATATATGCTAGCTGAAGTGAAGATGATTTTATTGAAGAGGTATT 295
QY 61 ValValMetGluAspSerProValGluAlaValSerThrProAsnThrProArgAsn 80
DB 296 GTTGTAATGGAGATGATTTCTCAGTGGAGGCTGTGAGACACACCTTAATCTCCCGAACC 355
QY 81 LeuAlaAlaTrpLysIleSerIleProTyrValAspPheGluAspProSerSerGlu 100
DB 356 CTTCGCTGCATGGAATATAGCATTCATATGATAGACTTTTTTGGAGATCCCTCTCTGAA 415
QY 101 ArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArgAsnAspArg 120
DB 416 AGCAAGCAGAAAAAAGAAAGAAATTCCTGTGTGTTTGTATTGATGATCTTGAAGAAATAGA 475
QY 121 ArgAlaValGlyHisGluProGluHisTrpSerValTyrArgArgTyrIleGluPheTyr 140
DB 476 AGAGCAGTGGACACAGAGCTGAACATTCCTGCTCTATAGAGATATCTTGAATCTCAT 535
QY 141 ValLeuGluSerLysLeuThrGluPheHisGlyValPheProAspAlaGlnLeuProSer 160
DB 536 GTACTTGAATCAAAATACAGAAATTTTCATGGTGCAATTCCTGTATGATGCCAGCTTCTCT 595
QY 161 LysArgIleGlyProLysAsnTyrGluPheLeuLysSerIleArgGluGluPheGln 180
DB 596 AAGAGGATCATTCGCCCAAAATATATGAATCTTAAAGTCAAGAGGAGGAGAGTCCAA 655
QY 181 GluTyrLeuGlnLysLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAsp 200
DB 656 GAATATCTACAGAAATCTTCGCAGCATCCAGACTCAGTAATAGTCAACTTCTGCGAGAC 715

```

```

201 PheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsn 220
DB 716 TTTCCTTCCCTTAATGGTGGGAAACACAAATTTCTTGATAAGATATACCAGATGTAAT 775
QY 221 LeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeu 240
DB 776 CTGGGAAATATATAAATCTGTCTCTGAAATCAATGAAAGAGAAAGAGTTCAGCATTTG 835
QY 241 GluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArg 260
DB 836 GAACCTTTTATCATGAATTTCAATTAATCTTGTAGTCTCCAAAGCCTTAACCAAGTAGA 895
QY 261 ProGluLeuThrIleLeuSerProThrSerGluAsnLysLysLeuPheAsnAspLeu 280
DB 896 CAGAACTGACCATTTCTGAGCCCTACTTCAGAAAACAAACAAGAGCTTTCAATGATCTG 955
QY 281 PheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGluAsnGlnAsnTyrPhe 300
DB 956 TTTTAAATTAATGCAAAACCGTCTGAAATATACAGAGAGAAAGCAAAATCAGAAATATT 1015
QY 301 MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal 320
DB 1016 ATGGAGGTGATGACTGTAGAGGAGTCTATGATTACCTGATGATGTAGAGCGGTAGTT 1075
QY 321 PheGlnValProAspTrpLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsn 340
DB 1076 TTCAGGTTCTGACTGGCTTCATCATCTCTTAATGGGAACCTCGAATCTCTTTTAAAC 1135
QY 341 ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGlu 360
DB 1136 ACCCTGGAAATGATATCTACTGATTACTTCTCAGTGTAAACTAGACAGCTATTTCAG 1195
QY 361 HisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGlu 380
DB 1196 CACCGTTTGGTCTCACTCAACACATCTCTCAGAGATGCTATATTTCTGTGAAACACTGAA 1255
QY 381 ProArgSerLeuGlnAspLysGlyAlaLysGlnThrPheGluGluMetMetAsn 400
DB 1256 CCTCGTCTCTCCAAAGATAAGCAAAAGAGCAAAACAGACTTTTGAAGAAATGATGAAT 1315
QY 401 TyrIleProAspLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArg 420
DB 1316 TACATTCAGATCTGTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGA 1375
QY 421 LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu 440
DB 1376 CTTCCTGTTTGTATGCTTACAGCAACCACTACTCAACAGCAGCTGACTTATGTTTATTG 1435
QY 441 AspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSer 460
DB 1436 GACATTTGTGATACAGGAACCTGTTTCCAGAGCTCAATAGGTACAAAAGGAAGTTACCTCT 1495
QY 461 ValThrSerTrpMet 465
DB 1496 GTGACATCTTGGATG 1510

RESULT 2
AY044865 2661 bp mRNA linear PRI 09-SEP-2001
LOCUS Homo sapiens sorting nexin 14 (SNX14) mRNA, complete cds.
DEFINITION
ACCESSION AY044865
VERSION AY044865.1 GI:15529063
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 2661)
AUTHORS Hong, W.
TITLE The complete coding region of SNX14
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 2661)
AUTHORS Hong, W.

```



ORGANISM	Homo sapiens	Query Match:	97.85%	Indels:	0
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	DB:	9	Gaps:	0
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	US-09-744-313A-1 (1-465) x BC005110 (1-3038)			
TITLE	1 (Bases 1 to 3038)				
JOURNAL	Submitted (26-MAR-2001) National Institutes of Health, Mammalian				
	Gene Collection (MGC), Cancer Genomics Office, National Cancer				
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,				
	USA				
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov				
COMMENT	Contact: MGC help desk				
	Email: gcapbs-remail.nih.gov				
	Tissue Procurement: ATCC				
	CDNA Library Preparation: Rubin Laboratory				
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)				
	BC Cancer Agency, Vancouver, BC, Canada				
	info@bcgsc.bc.ca				
	Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,				
	Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,				
	Leticia Hsiao, Martin Krywinski, Reta Kutsche, Oliver Lee, Soo				
	Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven				
	Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline				
	Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Scott,				
	Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,				
	George Yang, Scott Zuyderduyn, Marco Marra.				
	Clone distribution: MGC clone distribution information can be found				
	through the I.M.A.G.E. Consortium/ILNL at: http://image.llnl.gov				
	Series: IPAL Plate: 18 Row: m Column: 23				
	This clone was selected for full length sequencing because it				
	passed the following selection criteria: Hexamer frequency ORF				
	analysis.				
FEATURES	Location/Qualifiers				
source	1..3038				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="MGC:13217 IMAGE:3959086"				
	/tissue type="Placenta, choriocarcinoma"				
	/clone.Tib="NIH MGC 21"				
	/lab_host="DH10B-R"				
	/note="Vector: pOTB7"				
	189..2849				
	/codon_start=1				
	/product="Unknown (protein for MGC:13217)"				
	/protein_id="AAH05110.1"				
	/db_xref="GI:13477273"				
	/translation="MGQKLKQRLDLVGRERICQYPLFCFLLCLSAALLNRYIHI				
	LMIFSVAGVVFYCSIGPSLLPNFTPTIKYKQGLQELFPQGSQVAVGKVK				
	KRRPSLLLENYQWLDKISSKVDASLSEVDIPSIITKLLKAAKMHIEVIVKARQ				
	VNTEFLQAALEEVGPBLHVALRSRDELHVLKLTLLPPYILPKATDCRSLLTL				
	TRILSGSVFLPSLDPLADPTVNHLLIIFDDSPPEKATEPASPLVPFLQKAPEN				
	KPSVLKELQIRQQLPFRNFKQEGAVHVLQCLFVEEFNRIILRPSELNDE				
	MLSEELQIKYTKCLDESIDKIFDFIIVEIORIABGYIDVVKLQTRCLFEAY				
	EHVLSLENVTFPFCHSDYFQRLGAEPSRNSKLNTRTKRGSFGISRGSKI				
	KGVFSTWBMGMLPNYGVEGDQDFIEGIVMEDDSPVAEVSTPNTPRNLAAWKIS				
	LPYVDFEDPSSEKKEKRIIPVCIQVNRNDRAVGEPEHMSVRYLYEFLVLESK				
	LTFEGAPFDQPLPSKRIIGKPYEFLSKREBFQYELQKLLQPELSNSQLLADFLS				
	PNGGTQFLDLPLVNGLIKSVFGKLMKEKQHLFPFLMNFINSCESPKPSRP				
	ELTILSPSTSENNKKLFNLFKNANRAENTERKQNYFMWEVTVEGYILMYGVR				
	VQVFDWLEHLMGRILFKNTLNTYTLQCKLEQLQFQHRVLVSITLLRDAIFCE				
	NTEPSLDQKQKAGQKTEEMNNTIPDLVKGICBETKYSIRLLFDGLQOPVLNKLQ				
	TYVLDIVIQELFPPLNKVQEVTSWMM"				
BASE COUNT	949 a 564 c 642 g 883 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	0	Length:	3038		
Score:	455.00	Matches:	455		
Percent Similarity:	100.00%	Conservative:	0		
Best Local Similarity:	100.00%	Mismatches:	0		



Qy 351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeuLeu 370  
Db 2502 CAGTGTAAACTAGAACAGCTATTTCAGAGCAGCCGTTTGGTCTCACTCAACACTTCTC 2561  
Qy 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
Db 2562 AGAGATCTATATTCTGTGAACACATGAACCTCGCTCTCTCCAGATAGCAAAAGGA 2621  
Qy 391 AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
Db 2622 GCAAAACAGACTTTTGAGGAATGATGAATATACATCCAGATCTGTAGTCAAGTGATT 2681  
Qy 411 GlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProVal 430  
Db 2682 GGTGAAGAACCAAGATATGAAGCATCAGACTTCTGTGATGGCTTACAGCAACAGTA 2741  
Qy 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGlu 450  
Db 2742 CTCACAGAGAGCTGACTTATGTTTATTTGACATCTGTGACATCTTGGATG 2801  
Qy 451 LeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
Db 2802 CTCATTAAGGTACAAAGGAAGTTACCTCTGTGACATCTTGGATG 2846  
RESULT 4  
BC046520 2925 bp mRNA linear PRI 21-FEB-2003  
LOCUS Homo sapiens, Similar to sorting nexin 14, clone IMAGE:5267454,  
DEFINITION mRNA.  
ACCESSION BC046520  
VERSION BC046520.1 GI:28461364  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 2925)  
Strausberg, R.  
Direct Submission  
Submitted (03-FEB-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgapsb-remail.nih.gov](mailto:cgapsb-remail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [amgobcm.tmc.edu](mailto:amgobcm.tmc.edu)  
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulsegged, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAP Plate: 91 Row: A Column: 6.  
Location/Qualifiers  
1. 2925  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5267454"  
/tissue type="Testis"  
/clone lib="NIH MGC 97"  
/lab host="DH10B"  
/note="Vector: pBluescript"  
BASE COUNT 948 a 513 c 579 g 885 t

ORIGIN  
Alignment Scores:  
Pred. No.: 0 Length: 2925  
Score: 390.00 Matches: 454  
Percent Similarity: 99.56% Conservative: 0  
Best Local Similarity: 99.56% Mismatches: 1  
Query Match: 83.87% Indels: 2  
DB: 9 Gaps: 0  
US-09-744-313a-1 (1-465) x BC046520 (1-2925)  
Qy 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysLys 30  
Db 1381 AGGAACACACAGAAAGGGAGAAATCATTTGGAAATCAGAGAAATAGGTAGCAAAATATAA 1440  
Qy 31 GlyValPheLysSerThrThrMetGluGluAlaMetLeuProAsnTyrGlyValAlaGlu 50  
Db 1441 GGAGTATTCAAAGTACCACAAATGGAGGAGCTATGTTGCCATAATTATGTTAGTACGAA 1500  
Qy 51 GlyGluAspAspPheIleGluGluGlyValValMetGluAspAspSerProValGlu 70  
Db 1501 GGTGAGATGATTATTTGAAGAGGTATTGTTGTAATGGAAGATGATTTCCAGTGGAG 1560  
Qy 71 AlaValSerThrProAsnThrProArgAsnLeuAlaIleThrLysLysIleSerIleProTyr 90  
Db 1561 GGTGTGAGCACACCTAATATCTCCAGAAACCTTCTGTCATGCAAAATTAGCATTCATAT 1620  
Qy 91 ValAspPhePheGluAspProSerSerGluArgLysGlnLysLysGluArgIleProVal 110  
Db 1621 GTAGACTTTTTTGGAGGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAATTCCTGTG 1680  
Qy 111 PheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTyr 130  
Db 1681 TTTTGTATTGATTGTTGAAGAAATGATAGGAGAGCAGTTGGACAGGACCTTGACATGG 1740  
Qy 131 SerValTyrArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
Db 1741 TCTGCTATAGAAGATATCTTGAATTCATGCTACTTGAATCAAACTCAACAGAAATTCAT 1800  
Qy 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu 170  
Db 1801 GGTGCATTTCTGATGCCAGCTTCTTCTAAGAGGATCATTTGGCCCCCAAAATPATGAA 1860  
Qy 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190  
Db 1861 TTTCTTAAAGTCAAGAGGAGAGTTCCAGAAATATCTACAGAAACTTCTCGCAGCATCA 1920  
Qy 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
Db 1921 GAACGTAGTATAGTCAACTTCTGCGACACTTCTTCTCCCTTAATGTTGGGGAACACAA 1980  
Qy 211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysLeuSerValProGly 230  
Db 1981 TTTCTTGAATAGATATCTACAGATCTACAGATGTAATCTTGGGAAATATATAAAATCT 2040  
Qy 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSer 250  
Db 2041 AAACCTAATGAAGAGAAAGGTGAGCATTTGGAACCTTTTATCATGAATTTTCAATTTCT 2100  
Qy 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
Db 2101 TGTGAGTCTCCAAAGCCTTAACCAAGTAGACCAAGACTGACCATTTCTCAGCCCTACTTCA 2160  
Qy 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290  
Db 2161 GAACCAACAGAGAGCTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAAAT 2220  
Qy 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
Db 2221 ACAGAGAGAAAGCAAAATCAGAAATTTATTTATGAGAGGTGATGACTGTAGAGAGTCTAT 2280  
Qy 311 AspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHisLeu 330

2281 GATTACCTGATGATAGGACGGTAGTCTTCCAGGTTCTGACTGGCTTCATCATCTC 2340  
331 LeuMetGlyThrArgIleuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
2341 TTAATGGGAACCTCGAATCTCTTTAAACACACCCCTGGAAATGTACTGATTACTATCTT 2400  
351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuLeu 370  
2401 CAGTGTAACATAGACACACTATTTCCAGGACCCGTTGGTCTCCTCATTAACACTTCTC 2450  
371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
2461 AGAGATGCTATATCTTGTAACACACTGAACCTCGCTCTCTCCAAGATAAGCAAAAGGA 2520  
391 AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysII 410  
2521 GCACAAACAGACTTTTGAAGAATGATGAATG-CATTCCAGATCTGTTAGTCAAGTGAT 2579  
410 eGlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnProVa 430  
2580 TGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTTACAGCAACAGT 2639  
430 IleuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProG 450  
2640 ACTCAACAGCAGCTGACTTATGTTTATTTGACATTTGTATGATACAGGACTGTTTCCAGA 2699  
450 uLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
2700 GCTCAATAGGTACAAAGAGGATTTACCTGTGTGACATCTTGGATG 2745

RESULT 5  
AF121863 1593 bp mRNA linear PRI 14-SEP-2001  
LOCUS Homo sapiens sorting nexin 14 (SNX14) mRNA, partial cds.  
DEFINITION  
ACCESSION AF121863  
VERSION AF121863.1 GI:4689265

KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  
AUTHORS Teasdale, R.D., Locci, D., Houghton, P., Karlsson, L. and Gleeson, P.A.  
TITLE A large family of endosome-localized proteins related to sorting nexin 1

JOURNAL Biochem. J. 358 (Pt 1), 7-16 (2001)  
MEDLINE 21378165  
PubMed 11485546  
REFERENCE Teasdale, R.D., Gleeson, P.A. and Karlsson, L.  
AUTHORS Identification of eleven novel human sorting nexin molecules. A sub-group of the sorting nexin family is associated with the early endosomes

JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 1593)  
AUTHORS Teasdale, R.D., Gleeson, P.A. and Karlsson, L.  
TITLE Direct Submission  
JOURNAL Submitted (22-JAN-1999) The R.W. Johnson Pharmaceutical Research Institute, 3535 General Atomics Court, San Diego, CA 92121, USA

FEATURES  
source 1..1593  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
gene <1..1593  
/gene="SNX14"  
CDS <1..1125  
/gene="SNX14"  
/codon\_start=1  
/product="sorting nexin 14"  
/protein\_id="AAD27836.1"  
/db\_xref="GI:4689266"

/translation="DFFEDPSSERKEKRIPIVFCIDVNRDRAVGHPEHWSVYRR  
YLEFVLSEKITEFHGAPDAQLPSKRIIGPKNTEFLKSEEEFQYLOKLQHPELS  
NSQLADFLSPNGGETOFLDKILPDVNLGKIISKVPGKLMKEKQHLPEFLMNFINS  
ESPKEPRPLTILSPSENKKLNDLFPKNARERATERKONKONVEMVETVEGV  
IDYLMYGVVFPDNLHLHMGTRILFKNTLEMYTDYILCKLEQLFQHRHLVSLI  
TDLDAIFCENTEPRSDQKQKGFPEEMNITIPDLVVKICIGETIKYIESIRLFDG  
LQQPVINKQLTYVLIDIVIELFPELNKVKQEVTSVTSWM"

BASE COUNT 533 a 256 c 294 g 510 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 1593  
Score: 374.00 Matches: 374  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 80.43% Indels: 0  
DB: 9 Gaps: 0

US-09-744-313A-1 (1-465) x AF121863 (1-1593)

QY 92 AspPheGluAspProSerSerGluArgLysGluLysGluArgIleProValPhe 111  
Db 1 GACTTTTGTGAGATCCCTCTCTGAAAGGAAGGAGAAAAGAAAGAAATTCCTGTGTTT 60  
QY 112 CysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSer 131  
Db 61 TGTATTGATGTTGAAGAAATGATAGAGAGAGAGTTGGACAGAGCTGAACATTGGTCT 120  
QY 132 ValTyrArgArgTyrIleuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGly 151  
Db 121 GTCTATAGAAGATATCTTGAATTTCTATGTACTTGAATCAAAACTAACAGAAATTCATGT 180  
QY 152 AlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGluPhe 171  
Db 181 GCATTTTCTGATGCCAGCTTCTCTTAAAGAGAGATCATTTGGCCCAAAATATGAAATTC 240  
QY 172 LeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHisProGlu 191  
Db 241 TTAAGTCAAGAGAGGAGAGTTCCAGAAATATCTACAGAAACTTCTGAGCATCCAGAA 300  
QY 192 LeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGlnPhe 211  
Db 301 CTGAGTAATAGTCAACTTCTGGCAGACTTCTTCCCTTAATGGTGGGAAACACAATTT 360  
QY 212 LeuAspLysIleLeuProAspValLeuGlnLysIleIleLysSerValProGlyLys 231  
Db 361 CTTGATAAGATACCTACAGATGTAAATCTTGGGAAATATATAAATCTCTTCTCTGAAA 420  
QY 232 LeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCys 251  
Db 421 CTAAATGAAGAGAGAGAGTCAAGCATTTGGAACTTTTATCATGAATTTCAATTTCTGT 480  
QY 252 GluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGlu 271  
Db 481 GAGTCTCCAAGCCTAAACCAAGTAGACAGCACTTCCAGCACTTCTGAGCCTACTTCAGAA 540  
QY 272 AsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThr 291  
Db 541 AACACACAGAGAGCTTTTCAATGATCTGTTTAAAAATAATGCAAAACCGTCTGAAAAATA 600  
QY 292 GluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAsp 311  
Db 601 GAGAGAAAGCAAAATCAGAAATTTATTTATGGAGGTGATGACTGTAGAGAGAGTCTATGAT 660  
QY 312 TyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHisLeuLeu 331  
Db 661 TACCTGATGATGTAGAGCGGTAGTTTCCAGGTTTCTGACTGGCTTTCATCATCTCTTA 720  
QY 332 MetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGln 351  
Db 721 ATGGGACTCGAATCTCTTTTAAAAACACCCCTGGAATGTATATCTGATTACTATCTTCAG 780  
QY 352 CysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeuLeuArg 371

Db 781 TGTAACACTGAACAGCTATTTCAGGAGCACCGTTTGGTCTCACTCATACACTTCTCAGA 840  
Qy 372 AspAlailePheCysGluAsnThrGluProArgSerLeuGlnAspGlyGlnVal 391  
Db 841 GATGCTATATCTGTGAAACACTGAACCTCGCTCTCTCCAGATAGGAAAGAGCA 900  
Qy 392 LysGlnThrPheGluGluMetAsnThrIleProAspLeuValCysCysIleGly 411  
Db 901 AACACACTTTTCAAGAAATGATGAATTAATTCACAGATCTGTAGTCAAGTGTATTGCT 960  
Qy 412 GluGluThrLysThrGluSerIleArgLeuLeuPheAspGlyGlnGlnProValLeu 431  
Db 961 GAAGAAACCAAGATGAAGACATCAGACTTCTGTTGATGGTGTACAGCAACCACTACTC 1020  
Qy 432 AsnLysGlnLeuThrTyrrValLeuLeuAspIleValIleGluLeuPheProGluLeu 451  
Db 1021 AACAGCAGCTGACTTATGTTTATTGGACATGTGATACAGAACTGTTTCCAGAGCTC 1080  
Qy 452 AsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
Db 1081 AATAAGGTACAAAGAGGATGATCTCTGTGACATCTTGGATG 1122

RESULT 6  
AK026479 3576 bp mRNA linear PRI 29-SEP-2000  
LOCUS Homo sapiens cDNA: FLJ22826 fis, clone KAlA4022, highly similar to  
DEFINITION AF121863 Homo sapiens sorting nexin 14 (SNX14) mRNA.  
ACCESSION AK026479.1 GI:10439350  
VERSION oligo capping; fis (full insert sequence).  
KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (sites)  
AUTHORS Kawakami, T., Noguchi, S., Itoh, T., Shigetani, K., Senba, T.,  
Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Tanigami, A.,  
Fujiwara, T., Ono, T., Yamada, K., Fujii, Y., Ozaki, K., Hirao, M.,  
Ohmori, Y., Ota, T., Suzuki, Y., Oabayashi, M., Nishi, T., Shibahara, T.,  
Tanaka, T., Nakamura, Y., Iwagaki, T. and Sugano, S.  
NEDO human cDNA sequencing project  
Unpublished

TITLE 2 (bases 1 to 3576)  
REFERENCE Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
Shibahara, T., Tanaka, T. and Nakamura, Y.  
AUTHORS Direct Submission  
TITLE Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)

COMMENT NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

FEATURES Location/Qualifiers  
1..3576  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="KAlA4022"  
/tissue\_type="ileal mucosa"  
/clone\_lib="kaia"  
/note="cloning vector pWE18SFL3"  
1..3576  
misc\_feature  
/note="highly similar to AF121863 Homo sapiens sorting  
nexin 14 (SNX14) mRNA"

BASE COUNT 1181 a 633 c 735 g 1027 t  
ORIGIN

Alignment Scores:  
Pred. No.: 0 Length: 3576  
Score: 368.00 Matches: 403  
Percent Similarity: 99.51% Conservative: 0  
Best Local Similarity: 99.51% Mismatches: 1  
Query Match: 79.14% Indels: 2  
DB: 9 Gaps: 0  
US-09-744-313A-1 (1-465) x AK026479 (1-3576)  
Qy 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIleLys 30  
Db 2017 AGGAACACACAGAAAGGGAGGAATCATTTGGAAATCAGCAATAGGTAGCAAAATATAA 2076  
Qy 31 GlyValPheLysSerThrThrMetGluGlyValMetLeuProAsnThrGlyValAlaGlu 50  
Db 2077 GGAGTATTCAAAGTACACATGGAGGAGCTATGTTGGCTAATATATGTGTAGCTGAA 2136  
Qy 51 GlyGluAspPheIleGluGluGlyIleValValMetGluAspSerProValGlu 70  
Db 2137 GGTCAAGATGATTTTATTGAAGAGGTATTGTTGTAATGAAGATGATTCCTCCAGTGGAG 2196  
Qy 71 AlaValSerThrProAsnThrProArgAsnLeuAlaIleTyrLysIleSerIleProTyr 90  
Db 2197 GCTGTGAGCACACCTAATCTCTCCGAAACCTTCTGTCATGCAAAATTAGCATTTCCATAT 2256  
Qy 91 ValAspPheGluAspProSerSerGluArgLysGlyLysValGluArgIleProVal 110  
Db 2257 GTAGACTTTTGTAGGATCCCTCTCTGAAAGGAGGAGAAAGAAAGAAAGTCTCTGTG 2316  
Qy 111 PheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTyr 130  
Db 2317 TTTTGTATTGATTTGAAAGAAATGATAGAGACAGTGGACACGAGCCTGAACTGG 2376  
Qy 131 SerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
Db 2377 TCTGTCTATAGAGATATCTTGAATCTTATGATCTTCAATCAAACTAAACAGATTTTCAT 2436  
Qy 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu 170  
Db 2437 GGTGCATTTCTGTATGCCAGCTTCTTCTTAAGAGGATCATTTGCCCCCAAAATATGAA 2496  
Qy 171 PheLeuLysSerLysArgGluGluPheGlnTyrLeuGlnLysLeuLeuGlnHisPro 190  
Db 2497 TTTCTTAAAGTCAAAAGAGGAGAGAGTTCAGAAATATCTACAGAAACTTCTGCAGATCCA 2556  
Qy 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
Db 2557 GAACTGATATATGCTCACTTCTGCAGACTTTCTTTCCCTAATGTTGGGAAACACAA 2616  
Qy 211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValProGly 230  
Db 2617 TTTTCTTGATAAGATATACCAAGATGTAATCTTGGGAAATATATAAAATCTGTTCTCTGGA 2676  
Qy 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheLeuAsnSer 250  
Db 2677 AAATTAATGAAGAGAAAGGTCAGCATTTGGAACCTTTTATCATGAATTTTCAATTAATCT 2736  
Qy 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
Db 2737 TGTGAGTCTCCAAAGCCTTAACCAAGTAGACCAAGCTGACCATTTCTCAGCCCTACTTCA 2796  
Qy 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsn 290  
Db 2797 GAAACACAAAGAAAGCTTTTCAATGATCTGTTTAAAAAATATGCAAAACCGTGTCTGAAAT 2856  
Qy 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
Db 2857 ACAGAGAGAAAGCAAAATCAGAAATTAATTTATGAGGTGATGACTGTAGAGGAGTCTAT 2916  
Qy 311 AspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHisLeu 330



LOCUS AK000362 3145 bp mRNA linear PRI 22-FEB-2000  
 DEFINITION Homo sapiens cDNA FLJ20355 fis, clone HEP15804, highly similar to AF121863 Homo sapiens sorting nexin 14.  
 ACCESSION AK000362  
 VERSION AK000362.1 GI:7020397  
 KEYWORDS oligo capping; fis (full insert sequence).  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (sites)  
 AUTHORS Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T., Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.  
 TITLE NEDO human cDNA sequencing project  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 3145)  
 AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T., Shibahara, T., Tanaka, T. and Nakamura, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Department of Virology; Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan [E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416]  
 COMMENT NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan; cDNA full insert sequencing; Research Association for Biotechnology; cDNA library construction, 5'- & 3'-end one pass sequencing; Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).  
 FEATURES  
 source Location/Qualifiers  
 1..3145  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="HEP15804"  
 /cell\_line="HepG2"  
 /cell\_type="hepatoma"  
 /clone\_lib="HEP"  
 /notes="cloning vector pBS18SFL3"  
 misc\_feature 1..3145  
 /notes="highly similar to AF121863 Homo sapiens sorting nexin 14"  
 BASE COUNT 985 a 570 c 659 g 931 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 0 Length: 3145  
 Score: 330.00 Matches: 430  
 Percent Similarity: 99.54% Conservative: 0  
 Best Local Similarity: 99.54% Mismatches: 1  
 Query Match: 70.97% Indels: 2  
 DB: 9 Gaps: 0  
 US-09-744-313A-1 (1-465) x AK000362 (1-3145)  
 QY 35 SerThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGluGlyGluAspasp 54  
 DB 1673 AGTACCACATGAGGAGGAGCTATGTCCTAAATTATGGTGTAGCTGAAGGTGAAGATGAT 1732  
 QY 55 PheileGluGlyIleValValMetGluAspSerProValGluAlaValSerThr 74  
 DB 1733 TTATTGAAGAAGTATTGTTGTAATGGAAGATGATTCCTCAGTGGAGGCTGTGAGCACA 1792  
 QY 75 ProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIleProTyrValAspPhePhe 94  
 DB 1793 CCTAATACTCCCGAAACCTTGCTGCATGGAAAAATTAGCATTCCATATGATGATTTT 1852  
 QY 95 GluAspProSerGluArgLysGluLysLysGluArgIleProValPheCysIleAsp 114  
 DB 1853 GAGGATCCCTCTGTAAGGAGGAGGAAAAAGAAAGAAATTCCTGTGTTTGTATTGAT 1912

QY 115 ValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSerValTyrArg 134  
 DB 1913 GTTGAAGAAATGATAGAAGAGCAGTTGGACAGCAGCCTCAACATTTGGTCTGTCTATAGA 1972  
 QY 135 ArgTyrIleuGluPheTyrValIleuGluSerLysLeuThrGluPheHisGlyAlaPhePro 154  
 DB 1973 AGATATCTTGAATCTTATGTACTTGAATCAAACTAACAGAATTTTCATGGTGCAATTTCT 2032  
 QY 155 AspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLysSer 174  
 DB 2033 GATGCCAGCTTCTCTCTAAGAGATCATTTGGCCCCAAAANTATGAATCTTAAAGTCA 2092  
 QY 175 LysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHisProGluLeuSerAsn 194  
 DB 2093 AAGAGGGAAGAGTTCCAAGAATATCTACAGAACTTCTGCAGCATCCAGACTGAGTAAT 2152  
 QY 195 SerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLys 214  
 DB 2153 AGTCAACTCTGCGACACTTTCTTCCCTTAATGGTGGGAAACACAATTTCTTGTATAAG 2212  
 QY 215 IleleuProAspValAlaAsnLeuGlyLysIleIleLysSerValProGlyLysLeuMetLys 234  
 DB 2213 ATACTACCAGATGTAATCTTGGGAAATATATAAATCTGTTCTCGAAAACTAATGAAA 2272  
 QY 235 GluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerPro 254  
 DB 2273 GAGAAAGTCGAGATTTGGAACTTTATCATGAATTTCAATTAATTTCTGTGAGTCTCCA 2332  
 QY 255 LysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnLys 274  
 DB 2333 AAGCCTAAACCAAGTAGACCAGAACTGACCATTCTCAGCCCTACTTTCAGAAACAACAAG 2392  
 QY 275 LysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlnLeuThrGluLys 294  
 DB 2393 AAGCTTTTCAATGATCTGTTTAAAAAATAATGCAAAACCGTCTGAAAATATACAGAGAAG 2452  
 QY 295 GluAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMet 314  
 DB 2453 CAAATCAGAAATTTATTTATGGAGGTGATGACTGTAGAGAGTCTATGATTTACCTGATG 2512  
 QY 315 TyrValGlyArgValValPheGlnVal - ProAspTyrLeuHisIleLeuLeuMetGlyLys 334  
 DB 2513 TAATGAGAGCGGTAGTTTTCAGAT - TCCTGACTGGCTTCATCACTCTTATGGGAAC 2571  
 QY 334 rArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrIleuGlnCysLysLe 354  
 DB 2572 TCGAATCTCTTTAAAAACACCCCTGGAAATGTATCTGATTAATCTTCTCAGTGTAAC 2631  
 QY 354 uGluGlnPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaI 374  
 DB 2632 AGAACAGCTATTTTCAGAGCACCCTTTGGTCTCACTCATTAACACTTCTCAGAGATGCTAT 2691  
 QY 374 ePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnTh 394  
 DB 2692 ATTCTGTGAAACACTGAACCTCGCTCTCTCCAGATTAACAAAAGGAGACAAACAGAC 2751  
 QY 394 rPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluTh 414  
 DB 2752 TTTTGAAGAATGATGATTAATACATTCAGATCTGTTAGTCAAGTGTATTGGTGGGAAAC 2811  
 QY 414 rLysTyrGluSerIleArgLeuPheAspGlyLeuGlnGlnProValLeuAsnLysG 434  
 DB 2812 CAAGTATGAAGCATCAGACTTCTGTTTGTGGTGTACAGCAACCCAGTACTCTCAACAAGCA 2871  
 QY 434 nLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGluLeuAsnLysVa 454  
 DB 2872 GCTCACTATGTTTATTTATGGACATTTGGATACAGAACTGTTTCCAGAGCTCAATAAGGT 2931  
 QY 454 lGlnLysGluValThrSerValThrSerTrpMet 465  
 DB 2932 ACAAAGGAGTACCTCTGTGACATCTTGGATG 2965

RESULT 9	AK095380	3616 bp	mRNA	linear	PRI 15-JUL-2002
LOCUS	AK095380				
DEFINITION	Homo sapiens cDNA FLJ38061 fis, clone CTONG2014966, highly similar to SORTING NEXIN 14.				
ACCESSION	AK095380				
VERSION	AK095380.1 GI:21754626				
KEYWORDS	oligo capping, fis (full insert sequence).				
SOURCE	Homo sapiens				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M., Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Ozaki, K., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kiuchi, H., Kanda, K., Nagatsuma, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Ohshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahara, K., Masuho, Y., Nagai, K. and Isoigai, T.				
TITLE	NEDO human cDNA sequencing project				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 3616)				
AUTHORS	Isoigai, T. and Yamamoto, J.				
TITLE	Direct Submission				
JOURNAL	Submitted (04-JUL-2002) Takao Isoigai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan				
COMMENT	(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.				
FEATURES	Location/Qualifiers				
source	1..3616				
	/organism="Homo sapiens"				
	/mol_type="mRNA"				
	/db_xref="taxon:9606"				
	/clone="CTONG2014966"				
	/tissue_type="tongue, tumor tissue"				
	/clone_lib="CTONG2"				
	/note="cloning vector: pME18SPL3"				
BASE COUNT	1078 a 529 c 710 g 1199 t				
ORIGIN					
Alignment Scores:					
Pred. No.:	3,286-291	Length:	3616		
Score:	269.00	Matches:	453		
Percent Similarity:	98.91%	Conservative:	0		
Best Local Similarity:	98.91%	Mismatches:	2		
Query Match:	57.85%	Indels:	5		
DB:	9	Gaps:	0		
US-09-744-313A-1 (1-465) x AK095380 (1-3616)					
QY	11	ArgAsnThrGlnLysArgGlyGluSerPheGlyLeuSerArgIleGlySerLysIleLys	30		
Db	2080	AGCAACACACAGAAAAGGGAGATCAATTCGAATCAGCAGATAGGTAGCAAAATTA	2139		
QY	31	GlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGlu	50		
Db	2140	GGAGATTTCAAAGTACCAATGAGGAGGAGTATGTTGCCCTAAATATGGTGTAGCTGA	2199		
QY	51	GlyGluAspPheIleGluGlyIleValValMetGluAspSerProValGlu	70		
Db	2200	GGTGAGATGATTTATTGAAGAGGTATTTGTAAAGGAGATGATTCACAGTGAG	2259		
QY	71	AlaValSerThrProAsnThrProArgAsnLeuAlaTrpLysIleSerIleProTyr	90		

Db	2260	GCTGTGAGCACACCTAATACTCCCGAAACCTTGTGTCATGGAAATATTAGCATTCATAT	2319		
QY	91	ValAspPhePheGluAspProSerSerGluArgLysGlu-LysLysGluArgIleProVa	110		
Db	2320	GTAGACTTTTTTGAGGATCCCTCTCTGAAGAGAGAGAAAAGAAAGAAATCTCTGT	2379		
QY	110	lPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisIst	130		
Db	2380	CTTTTGTATTGATGTTGAAGAAATGATAGAGAGCAGTTGGACACGAGCTGAACATTG	2439		
QY	130	pSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHi	150		
Db	2440	CTCTGTCTATAGAGATATCTTGAATCTATGTACTTGAATCAAACTAAACAGAAATTC	2499		
QY	150	sGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGl	170		
Db	2500	TGGTGCAATTTCTGATGCCAGCTTCCTCTPAAGAGGATCATTTGCCCCCAAAATTA	2559		
QY	170	uPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHisPr	190		
Db	2560	ATTCTTAAGTCAAGAGGAGAGTTCACAGAAATATCTACAGAACTTCTGCAGCATCC	2619		
QY	190	oGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGl	210		
Db	2620	AGAAGCTGAGTAATAGTCAACTTCTGGCAGACTTTCTTCCCTAATGCTGGGAAAC	2679		
QY	210	nPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValProGl	230		
Db	2680	ATTTCTTGAATAGTACTACAGATGTAAATCTTTGGGAAATATATAAAATCTGTCTCTGG	2739		
QY	230	YValLeuMetLysGluLysGlyGlnHisLeuGluPro-PheIleMetAsnPheIleAsn	250		
Db	2740	AAACTATATGAAGAGAAAGGTGACATTTGGAAT-TTTTATCATGAATTTCAATTA	2798		
QY	250	erCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrS	270		
Db	2799	CTTGTGAGTCTCCAAAGCCTTAAACCAAGTAGACCAAGAACTGACCACTTCTCAGCC	2858		
QY	270	erGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu	290		
Db	2859	CAGAAACAAACAAAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTGTGAA	2918		
QY	290	snThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal	310		
Db	2919	ATACAGAGAGAAAGCAAAATCAGAAATTTATTCGAGGTGATGACTGTAGAGAGTCT	2978		
QY	310	YzAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHisIst	330		
Db	2979	ATGATTTACCTGATGATGTAGTAGCAGGTAGTTTCCAGGTTCTCTGCTGCTTCATC	3038		
QY	330	euLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTrL	350		
Db	3039	TCTTAATGGAACTCGAATCTCTTTAAACACACCGCTGAAATGTATATCTGATTACT	3098		
QY	350	euGlnCysLysLeuGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuL	370		
Db	3099	TTCAAGTGTAACTAGAACAGCTATTTCAGGAGCAGCGTTTGGTCTCCTCATTAAC	3158		
QY	370	euArgAspAlaIlePheCysGluAsnThrGlu-ProArgSerLeuGlnAspLysGlnLys	389		
Db	3159	TCAGAGATGCTATATTCTGTGAAACACTCA-ACCTCGCTCTCTCAAGATAGCAAAA	3217		
QY	390	GlyValLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys	409		
Db	3218	GGAGCAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTAGTCAAGTGT	3277		
QY	410	IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnPro	429		
Db	3278	ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGTGGTGGCTTACAG	3337		
QY	430	ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro	449		
Db	3338	GTACTCAAGAGCAGCTGACTTATGTTTTTATTTGGACATTTGTGTATCAGGAATG	3397		





```

337 LeuPheLysAnthrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGln 356
121 CTCTTAAACACACCTGGAAATGATATACCTGATCTTCTCAGTCTAACTAGAACAG 180
357 LeuPheGlnGluHisArgLeuValSerLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 376
181 CTATTTTCAGAGACACCGTTTGGTCTCACTCATACACTTCTCAGAGATGCTATATTCTGT 240
377 GluAnThrGluProArgSerLeuGlnAspLysGlnLysGlnLysGlnLysGlnLys 396
241 GAAACACACTGAACCTCGCTCTCTCAAGATAGCAAAAGAGCAAAACAGACTTTTGAA 300
397 GluMetMetAnthrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyr 416
301 GAATGATGATTTACATTCAGACTCTGTAGTCAAGTGTATTGTTGAAGAACCAAGTAT 360
417 GluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuLeuLeuLysGlnLeuThr 436
361 GAAAGCATCAGACTTCTGTTGATGGCTTACAGCAACCACTACTCAACAGCAGCTGACT 420
437 Tyr 437
421 TAT 423

RESULT 12
BD026500
LOCUS BD026500 358 bp DNA linear PAT 27-AUG-2002
DEFINITION Sequence tag and encoded human protein.
ACCESSION BD026500
VERSION BD026500.1 GI:22567723
KEYWORDS JP 2001269182-A/2746.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 358)
EDWARDS, J.B.D.M., Duclair, E. and Jordan, J.Y.
TITLE Sequence tag and encoded human protein
JOURNAL Patent: JP 2001269182-A 2746 02-OCT-2001;
GENSET
COMMENT OS Homo sapiens (human)
PN JP 2001269182-A/2746
PD 02-OCT-2001
PF 24-FEB-2000 JP 2000118773
PR 26-FEB-1999 US 60/122487
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC
C12N5/10,
PC C12P21/02, C12P21/08, C12Q1/68//G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40
CC
I PH KEY Location/Qualifiers
FT CDS Location/Qualifiers
1..358
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 125 a 63 c 59 g 101 t
ORIGIN
Alignment Scores:
Pred. No.: 1.16e-123 Length: 358
Score: 119.00 Matches: 119
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 25.59% Indels: 0
DB: 6 Gaps: 0

US-09-744-313a-1 (1-465) x BD026500 (1-358)

```

```

216 LeuProAspValAnLeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGlu 235
2 CTACCAGATGTAATCTTGGGAAATATATAAATCTGTCTCTGAAACATAATGAAAGAG 61
236 LysGlyGlnHisLeuGluProPheIleMetAsnPheIleLeuSerCysGluSerProLys 255
62 AAAGTCAGCATTTGGAAACCTTTTATCATGATTTTCAATTAATTTCTGTGAGTCTCCAAAG 121
256 ProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLys 275
122 CCTAAACCAAGTAGACCAAGAACTGACCATTTCTAGCCCTACTTCAGAAAAACACCAAGAAG 181
276 LeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAnThrGluArgLysGln 295
182 CTTTTCATGATCTGTGTTTAAATAATGCAACCGTGTGAAATACAGAGAGAAGCAA 241
296 ArgGlnAnThrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyr 315
242 AATCAGATTAATTTTATGGAGGTGATGACTGTAGAAGGAGTCTATGATTACCTGATGAT 301
316 ValGlyArgValAlaPheGlnValProAspTyrLeuHisLeuLeuMetGlyThr 334
302 GTAGGACGGTAGTTTTCAGGTTCCAGTGGCTTCATCATCTCTTAATGGGAACC 358

RESULT 13
BC043328
LOCUS BC043328 1782 bp mRNA linear ROD 16-APR-2003
DEFINITION Mus musculus RIKEN cDNA C330035N2 gene, mRNA (cdna clone MGC:49424
IMAGE:540157), complete cds.
ACCESSION BC043328
VERSION BC043328.1 GI:27694048
KEYWORDS MGC.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1782)
STRAUSBERG, R.L., Feingold, E.A., Grouse, L.H., Derge, J.G.,
Klausner, R.D., Collins, P.S., Wagner, L., Sherman, C.M., Schuler, G.D.,
Altshul, S.P., Zeeberg, B., Buetow, K.H., Schaefer, C.F., Bhat, N.K.,
Hopkins, R.F., Jordan, H., Moore, T., Max, S.I., Wang, J., Hsieh, F.,
Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,
Scheetz, T.E., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,
Carninci, P., Prange, C., Raha, S.S., Loquellano, N.A., Peters, G.J.,
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,
McKernan, K.J., Malek, J.A., Gunaratne, P.H., Richards, S.,
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,
Villalón, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,
Fahey, J., Helton, E., Kettman, M., Madan, A.C., Rodrigues, S.,
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,
Boutard, G.G., Blakeley, R.W., Touchman, J.W., Green, E.D.,
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,
Butterfield, Y.S., Krzywinski, M.I., Skalska, U., Smalley, D.E.,
Schmerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.
TITLE Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE 22388257
PUBMED 12477932
REFERENCE 2 (bases 1 to 1782)
AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (09-JAN-2003) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgapps@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```

DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
info@bcgsc.bc.ca

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield, Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin, Letlicia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo Len Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline Schein, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Stott, Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy, George Yang, Scott Zuyverduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAX Plate: 86 Row: p Column: 11  
 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

## FEATURES

## Source

Location/Qualifiers

1. 1782

/organism="Mus musculus"

/mol type="mrna"

```
/db_xref="taxon:10090"
```

```
/usr_xlci= /usr_xlci:10050  
/clone="MGC:49424 IMAGE:5400157"
```

```

/tissue type="Eye retina mouse strain C57BL/6"
/clone="MGC:43424 IMAGE:340013"

```

```
/class_type="byte, rectm
/columns Tib="INTH MSC 94"
```

```
clone_lib="NLH_MGC_94"
lib_how="top_q11"
```

```

/ tab_host="DHTUB"
/ host="192.168.1.100"

```

```
/note="Ve
```

1. 1782

/gene="C330035N22Rik"

```
/note="synonym: B830022K16"
```

**/db\_xref="LocusID:244962"**

$$/db\_xref="N$$

228. .1616

/codon\_start=1

/product="C330035N22R1k pro

/protein\_id="AAH43328.1"

```
/db xref="GI:27694049"
```

/db\_xref="locusID:244962"  
/translation="MRCLIPFRSTQKGBSGISRIQSKIKGVFKSTMEGAVLPNG  
VAAGSDDFIEGIVMDDSEVAVSTPNTPRNLAANKISIPYDFPDPSSRKKK  
IRIPYFCDIVDERAVAGHPSEVVRVLYFVLELSPKTHGTFPDAQPSRKI  
IGPNYFPLKSRSEFQVQLVQHLEPNSQLADPLSPNGGTEGTDKILDPVNL  
GKLIKSPGKLMKEGQHLPEPIMPSPINSCSPKPSRPELILFPTSENKKLFND  
LPFNNAARNTERKQNYFMVEVTVGVYVILVYGVRFQVPMWLHLMGTIRL  
LIPNNTDYYQLQSQLESLYSITLLERDAIFCENTPSRLDQKQKQATF  
EEMWYIPDLIVKCIIGETKYESTIRLLPDGQQQVPLNKQLTYVLLDVIQELFPELNK  
VQKATSTSTSW"

Wavelength, nm	339	380	496
100	0.00	0.00	0.00
200	0.00	0.00	0.00
300	0.00	0.00	0.00
400	0.00	0.00	0.00
500	0.00	0.00	0.00
600	0.00	0.00	0.00
700	0.00	0.00	0.00
800	0.00	0.00	0.00
900	0.00	0.00	0.00
1000	0.00	0.00	0.00

**BASCO OPTICAL**

# ORIGIN

Alignment Scores:		
Pred. No.:	8.26e-112	Length:
Score:	109.00	Matches:
Percent Similarity:	97.34%	Conservative:
Best Local Similarity:	97.34%	Mismatches:
Query Match:	23.44%	Indels:
DB:	10	Gaps:
		1782

US-09-744-313A-1 (1-465) x BC043328 (1-1782)

13	ThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerIlySerIlyGlyVal	32
QY		
255	ACACAAAAGAGGAGGAATCGTTTGGAAATCAGCAGAAATAGTAGCAAAATTAAGCGGTA	314
DB		
33	PheLysSerThrThrMetGluGlyAlaMet-LeuProAsnTyGlyValalacIuGlyG1	52
QY		
315	TTCAAGAGTACGACAAATGAGGAGCTGT-CTCGCTAAATTACGGGGTGGCTCAAGCGGA	373
DB		
52	uAspAspPheIleGluGlyIleValValMetGluAspAspSerProValGluIaVa	72
QY		
374	AGATGACTTATTGAAGAAGGGATGTGGTAATGGAGGATGACTCTCCAGTAGAAGCTGT	433
DB		

Db 1509 CTCAACAGCAGCTGACCTATGTTCTGTCGACATTGTGATACAGAGCTGTTTCCAGAG 1568

QY 451 LeuAenLysValGlnLysGlu 457

DB 1569 CTAATAAGGTACAAAGGAA 1589

RESULT 14

AL589666/c AL589666 129010 bp DNA linear PRI 12-MAY-2001

LOCUS Human DNA sequence from clone RP11-321N4 on chromosome 6, complete

DEFINITION

ACCESSION AL589666

VERSION AL589666.5 GI:14041764

KEYWORDS HTG.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 129010)

AUTHORS Ramsay,H.

TITLE Direct Submission

JOURNAL Submitted (12-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk

COMMENT requests: clonerequest@sanger.ac.uk

On May 15, 2001 this sequence version replaced gi:11751565.

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one W13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C\_elegans/wormpep

This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr6

RP11-321N4 is from the library RP11-11.2 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-321N4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true left end of clone RP11-3017 is at 128911 in this sequence. The true right end of clone RP11-30P6 is at 100 in this sequence.

Location/Qualifiers

1..129010

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/chromosome="6"

/clone="RP11-321N4"

/clone\_lib="RP11-11.2"

2..192

/note="MER21B repeat: matches 3. .207 of consensus"

240..909

/note="MER21B repeat: matches 30. .715 of consensus"

1593..1682

/note="MIR repeat: matches 35. .145 of consensus"

2030..2102

/note="L2 repeat: matches 2681. .2750 of consensus"

repeat\_region

repeat\_region

repeat\_region

repeat\_region

3459..3574

/note="FLAM C repeat: matches 1. .117 of consensus"

5635..5670

/note="L1ME repeat: matches 683. .718 of consensus"

6887..6982

/note="L2 repeat: matches 2616. .2704 of consensus"

8284..8356

/note="L2 repeat: matches 2639. .2710 of consensus"

8380..8405

/note="L3 copies 2 mer tt 100% conserved"

8440..8601

/note="MIR repeat: matches 1. .169 of consensus"

8825..9360

/note="L2 repeat: matches 1686. .2301 of consensus"

9423..9592

/note="L2 repeat: matches 2560. .2727 of consensus"

12088..12175

/note="44 copies 2 mer at 73% conserved"

12110..12177

/note="17 copies 4 mer atat 79% conserved"

12790..12905

/note="L2 repeat: matches 2628. .2750 of consensus"

13672..13923

/note="L2 repeat: matches 2462. .2725 of consensus"

14774..15053

/note="AluSq repeat: matches 1. .289 of consensus"

15966..16031

/note="6 copies 11 mer 77% conserved"

17304..17359

/note="MIR repeat: matches 196. .251 of consensus"

17554..17682

/note="MIR repeat: matches 108. .228 of consensus"

18366..18701

/note="MER2 repeat: matches 1. .345 of consensus"

18959..19248

/note="AluGo repeat: matches 1. .284 of consensus"

20386..20592

/note="L2 repeat: matches 70. .288 of consensus"

20609..20778

/note="UTR28 repeat: matches 848. .1013 of consensus"

20747..21117

/note="UTR1 repeat: matches 295. .680 of consensus"

21163..21226

/note="UTR28 repeat: matches 297. .361 of consensus"

21187..21359

/note="UTR1 repeat: matches 16. .198 of consensus"

22191..22355

/note="FRAM repeat: matches 4. .163 of consensus"

22844..22991

/note="MIR repeat: matches 48. .202 of consensus"

23155..23190

/note="18 copies 2 mer ga 86% conserved"

24390..24440

/note="L2 repeat: matches 2657. .2708 of consensus"

25082..25109

/note="7 copies 4 mer aac 100% conserved"

25438..25744

/note="L2 repeat: matches 2427. .2744 of consensus"

25751..25917

/note="MIR repeat: matches 95. .262 of consensus"

25964..26307

/note="L2 repeat: matches 1986. .2366 of consensus"

26526..26673

/note="FLAM C repeat: matches 1. .142 of consensus"

26682..26737

/note="14 copies 4 mer agga 76% conserved"

28208..28394

/note="MIR repeat: matches 12. .196 of consensus"

29052..29250

/note="MIR repeat: matches 48. .251 of consensus"

30082..30170

/note="MIR repeat: matches 50. .145 of consensus"

30197..31090

repeat\_region

```

/notes="VER11D repeat: matches 1. .897 of consensus"
31921. .32067
/notes="FRAM repeat: matches -2. .151 of consensus"
32545. .32678
/notes="L2 repeat: matches 2578. .2708 of consensus"
32860. .32895
/notes="9 copies 4 mer tttt 83% conserved"
32897. .33193
/notes="AluX repeat: matches 1. .296 of consensus"
35445. .35478
/notes="MIR repeat: matches 84. .118 of consensus"
37442. .37598
/notes="MER5B repeat: matches 1. .173 of consensus"
38659. .38860
/notes="MIR repeat: matches 15. .236 of consensus"
40552. .40911
/notes="THE1B repeat: matches 1. .364 of consensus"
40912. .42540
/notes="THE1B-INTERNAL repeat: matches 1. .1580 of consensus"
42541. .42899
/notes="THE1B repeat: matches 1. .364 of consensus"
42900. .43231
/notes="MT1A1 repeat: matches 4. .349 of consensus"
43407. .43851
/notes="MT1C repeat: matches 9. .466 of consensus"
44641. .45020
/notes="MSTA repeat: matches 1. .382 of consensus"
45689. .45988
/notes="AluSg repeat: matches 1. .302 of consensus"
46220. .46351
/notes="Tigger3b repeat: matches 567. .698 of consensus"
46358. .46609
/notes="AluJb repeat: matches 40. .290 of consensus"
46623. .46758
/notes="AluJo/FRAM repeat: matches 170. .301 of consensus"
46759. .47229
/notes="Tigger3b repeat: matches 2. .540 of consensus"
47230. .47442
/notes="MER44A repeat: matches 85. .327 of consensus"
47715. .48005
/notes="L1PA10 repeat: matches 5864. .6153 of consensus"
48577. .48973
/notes="MT1B repeat: matches 1. .350 of consensus"
50633. .50925
/notes="AluY repeat: matches 3. .295 of consensus"
53828. .55056
/notes="L1MC1 repeat: matches 5023. .6332 of consensus"
55092. .56081
/notes="L1M4 repeat: matches 3778. .4798 of consensus"
56074. .56651
/notes="L1MCA repeat: matches 579. .1150 of consensus"
56652. .56952
/notes="AluX repeat: matches 1. .301 of consensus"
56953. .57277
/notes="L1MCA repeat: matches 253. .579 of consensus"
57280. .57562
/notes="L2 repeat: matches 1590. .1886 of consensus"
57699. .57766
/notes="L2 repeat: matches 2639. .2710 of consensus"
59295. .59599
/notes="AluX repeat: matches 1. .307 of consensus"
59749. .59811

```

## Alignment Scores:

```

Pred. No.: 5 6e-63 Length: 129010
Score: 67.00 Matches: 67
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 14.41% Indels: 0
DB: Gaps: 0

```

US-09-744-313a-1 (1-465) x AL589666 (1-129010)

```

Qy 56 ILeGluGluGlyTleValValMetGluAspSerProValGluAlaValSerThrPro 75
Db 78112 ATTGAAGAAGGATTGTTGTAATGGAAGATGATCTCCAGTGAGGCTGTGACACCT 78053
Qy 76 AsnThrProArgAsnLeuAlaLaTPlYsIleSerIleProTyrValAspPhePheGlu 95
Db 78052 AATATCCCCGAACCTTGCTGATGGAATAATGATTCATATCCATATGATAGATTTTGTAG 77993
Qy 96 AspProSerSerGluArgLysGluLysLysGluArgIleProValPheCysIleAspVal 115
Db 77992 GATCCCTCTCTGAAAGGAGGAGAAAAAGAAAGAAATTCCTGTGTTTGTATGTATGTT 77933
Qy 116 GluArgAsnAspArgAla 122
Db 77932 GAAAGAATGATAGAGCA 77912
RESULT 15
AC116713/c
LOCUS AC116713 278898 bp DNA linear HTG 09-JUN-2003
DEFINITION Mus musculus clone RP23-118M23, *** SEQUENCING IN PROGRESS ***, 7
ordered pieces.
ACCESSION AC116713
VERSION AC116713.4 GI:131544102
KEYWORDS HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 278898)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Mus musculus, clone RP23-118M23
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 278898)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campiano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,J., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lechoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severi,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 278898)
REFERENCE
AUTHORS Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,I., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrum,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 02:44:36 ; Search time 85 Seconds

(without alignments)

2414.625 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 2437

Sequence: 1 MYLHCLIFRNTQKRGESF.....ELFPELNKYQKEVTSWIM 465

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09744313/runat 30012004 113536 4209/app query.fasta\_1.647  
-DB=Issued Patents NA -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313.@CIGN 1.1.56.@runat 30012004 113536 4209 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEBLOCK -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*

- 1: /cgn2\_6/ptodata/2/ina/5A COMB.seq.\*
- 2: /cgn2\_6/ptodata/2/ina/5B COMB.seq.\*
- 3: /cgn2\_6/ptodata/2/ina/6A COMB.seq.\*
- 4: /cgn2\_6/ptodata/2/ina/6B COMB.seq.\*
- 5: /cgn2\_6/ptodata/2/ina/PTUS COMB.seq.\*
- 6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2386	97.9	1716	4	US-09-620-312D-922
2	2057.5	84.4	1551	4	US-09-620-312D-923
3	136.5	5.6	1497	4	US-09-620-312D-1021
4	128	5.3	1974	1	US-08-625-322-1
5	116	4.8	3747	2	US-09-080-897-1
6	116	4.8	3747	3	US-09-323-735-1
7	113.5	4.7	5822	3	US-08-899-595-4
8	113.5	4.7	5822	3	US-08-899-595-5
9	110	4.5	580073	4	US-08-545-528D-1
10	109.5	4.5	580073	4	US-08-545-528D-1
11	108.5	4.5	9409	4	US-08-961-527-161
12	108	4.4	3621	4	US-09-635-872A-21

13	108	4.4	3621	4	US-09-636-077A-21
14	108	4.4	3783	4	US-09-635-872A-20
15	108	4.4	3783	4	US-09-636-077A-20
16	105	4.3	4065	4	US-09-016-434-1105
17	105	4.3	4739	3	US-08-585-871-1
18	103	4.2	1353	1	US-08-625-322-3
19	103	4.2	5597	4	US-09-635-872A-4
20	103	4.2	5597	4	US-09-636-077A-4
21	103	4.2	1664976	4	US-08-916-421B-1
22	102.5	4.2	3095	6	5231168-1
23	102.5	4.2	4378	2	US-09-080-897-3
24	102.5	4.2	4378	3	US-09-323-735-3
25	102.5	4.2	4399	3	US-08-899-595-2
26	101.5	4.2	1330	2	US-09-036-582-33
27	101.5	4.2	8503	4	US-09-620-312D-130
28	101	4.1	5893	1	US-08-592-126-54
29	101	4.1	5893	2	US-08-687-080-44
30	101	4.1	5893	4	US-09-168-595-54
31	100.5	4.1	6519	1	US-08-588-985-1
32	100.5	4.1	6519	1	US-08-971-988-1
33	100	4.1	15223	3	US-08-962-690-12
34	100	4.1	15223	2	US-08-892-403A-1
35	100	4.1	15223	3	US-08-720-132-1
36	99.5	4.1	4100	4	US-09-620-312D-81
37	99	4.1	977	3	US-07-861-458C-97
38	99	4.1	2027	4	US-09-484-970B-98
39	98.5	4.0	424	4	US-09-397-787-133
40	98.5	4.0	3309	4	US-09-510-543-20
41	98.5	4.0	4982	3	US-08-699-103B-1
42	98.5	4.0	4982	4	US-09-229-059-1
43	98.5	4.0	4982	4	US-09-628-133-1
44	98.5	4.0	5878	4	US-09-510-543-18
45	98.5	4.0	1664976	4	US-08-916-421B-1

#### ALIGNMENTS

#### RESULT 1

US-09-620-312D-922  
; Sequence 922, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenchua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Reiyao  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Dmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: DT\_FL\_genes Version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716  
; TYPE: DNA

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (71)..(1543)  
US-09-620-312D-922

Alignment Scores:  
Pred. No.: 5,99e-281 Length: 1716  
Score: 2386.00 Matches: 456  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 97.91% Indels: 0  
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-620-312D-922 (1-1716)

QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerIleVal 29  
DB 173 TTTCCGACACACAGAAAGGGAGATCATTTGGATCAGCAGAAATAGGTAGCAAAATT 232  
QY 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
DB 233 AAAGGAGTATTCAAAGTACCAATGGAGGAGCTATGTTGCCTAAATTATGGTGTAGCT 292  
QY 50 GluGlyGluAspPheIleGluGlyIleValValMetGluAspSerProVal 69  
DB 293 GAAGGTGAGATGATTTTATTGAGAGGTATTTGTTGTAATGGAAGATGATTCCTCAAGTG 352  
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTTPLeIleSerIlePro 89  
DB 353 GAGGCTGTGACACACCTTAATCTCCGAAACCTTGTGCTGATGGAATAATTAGCATTTCCA 412  
QY 90 TyrValAspPheGluAspProSerGluArgLysGluLysGluArgIlePro 109  
DB 413 TATGTAGACTTTTGTGAGGATCCCTCTGAAAGGAGGAGAAAGAAAGAAATTCCT 472  
QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
DB 473 GTGTTTGTATTGATGTTGAAAGAAATGATAGAGAGCATGTTGGACACGAGCCTGAACAT 532  
QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
DB 533 TGTCTGTCTATAGAGATATCTTGAATCTATGACTTTGATCTGAACTCAAACTAACAGAAATT 592  
QY 150 HisGlyValPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
DB 593 CATGGTGATTTCTCTGATGCCAGCTTCTCTTAAGAGATCATTTGGCCCCCAAAATTTAT 652  
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHis 189  
DB 653 GAATTCCTTAAAGTCAAGAGGAGAGAGTTCCAAAGAAATATCTACAGAAACTTCTGACGAT 712  
QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGluThr 209  
DB 713 CCAGAACTAGTAATAGTCAACTCTGGCAGACTTCTTCTCCCTTAATGGTGGGAAACA 772  
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro 229  
DB 773 CAATTTCTTGATAGATATCTCCAGATGTAATCTTGGGAAATATTAATCTCTTCT 832  
QY 230 GlyLeuLeuMetLysGlyLysGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
DB 833 GGAATACTAATGAAGAGAAAGGTGAGCATTTGGAACTTTTATCATGAATTTTCAATTAAT 892  
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
DB 893 TCTTGTGAGTCTCAAGGCTTAACCAAGTAGACCACTGACCACTTCTCAGCCCTACT 952  
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGlu 289  
DB 953 TCAGAAACCAACAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAA 1012  
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309

DB 1013 AATACAGAGAGAAAGCAAAATCAGAAATATTTATGGAGGTGATGACTGTAGAGGAGCTC 1072  
QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrPheHis 329  
DB 1073 TATGATTACCTCATGTATGTAGGACGGTAGTTTCCAGGTTCTGACTGGCTTCATCAT 1132  
QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
DB 1133 CTCCTTAATGGGAACTCGAATCTCTTTAAACACCCCTGGAAATGTATATCTGATTACTAT 1192  
QY 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369  
DB 1193 CTTCACTGTAAACTAGACAGCTATTTCCAGGAGCACCGTTTGGTCTCACTCATAACACTT 1252  
QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
DB 1253 CTCAGAGATGCTATATTTCTGTGAAAAACATCAACCTCGCTCTCTCAAGATAAGCAAAA 1312  
QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
DB 1313 GGAGCAAAAACAGACTTTTGAAGAAATGATGAATTCATTCAGATCTGTTAGTCAAGTGT 1372  
QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
DB 1373 ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCAACCA 1432  
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
DB 1433 GTACTCAACAAGCAGCTGACTTATGTTTATTTGGCACTTGTGATACAGGAACTGTTTCCA 1492  
QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
DB 1493 GAGCTCAATAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 1540

## RESULT 2

US-09-620-312D-923  
; Sequence 923, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunding  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghaast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL genes Version 1.0  
; SEQ ID NO 923  
; LENGTH: 1551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:

NAME/KEY: CDS  
LOCATION: (71)..(1378)  
US-09-620-312D-923

Alignment Scores:  
Pred. No.: 6,28e-241 Length: 1551  
Score: 2057.50 Matches: 401  
Percent Similarity: 87.94% Conservative: 0  
Best Local Similarity: 87.94% Mismatches: 0  
Query Match: 84.43% Indels: 55  
DB: 4 Gaps: 1

US-09-744-313a-1 (1-465) x US-09-620-312D-923 (1-1551)

QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyLeuSerArgGlySerLysLeu 29  
DB 173 TTTCCGACACACAGAAAGGGAGATCATTTGGAATCAGCAGATAGGTAGCAAAAT 232  
QY 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
DB 233 AAAGGATATTCAAAGTACCAATCGAGGGAGCTATGTTGCTAATTATGTTAGCT 292  
QY 50 GluGlyGluAspPheIleGluGlyValValMetGluAspSerProVal 69  
DB 293 GAAGGTGAAGATGATTTTATTAAGAAGGATTTGTTAATGGAAGATGATTTCCAGTG 352  
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaLalaTyrLysIleSerIlePro 89  
DB 353 GAGGCTGTGACACACACCTTAATACCTCCGCAACCTTCCTCATGGAATAATAGCATTTCCA 412  
QY 90 TyrValAspPheGluAspProSerSerGluArgLysGluLysLysGluArgIlePro 109  
DB 413 TATGTAGATTTTTTTGAGATCCCTCTCTCTGAAGAGAGAGAGAGAGAGAGAGATTCCT 472  
QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
DB 473 GTGTTTCTATGATTTGTTGAAGAAATGATAGAGAGAGAGAGAGAGAGAGAGATTTAT 532  
QY 130 TrpSerValTyrArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
DB 533 TGGTCTGTCTATAGAGATATCTTGAATCTTATGATCTTGAATCAAACTTAACAGAAATTT 592  
QY 150 HisGlyAlaPheProAspAlaGluLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
DB 593 CATGTTGATTTCTCTGATGCTCCCTCTTAAAGAGATCATTTGGCCCCCAAAATTTAT 652  
QY 170 GluPheLeuLysSerLysArgGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis 189  
DB 653 GAATTCCTTAAGTCAAGAGAGAGAGAGTTCAGAGATATCTACAGAACTTCTGCAGCAT 712  
QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
DB 713 CCAGAACTGAGTAAATAGTCAACTTCTGCGAGACTTTCTTTCCCTAATGTTGGGAAACA 772  
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro 229  
DB 773 CAATTTCTTTGATAGATACTACAGATGTAATCTTGGGAAATTAATAAATCTGTTCCT 832  
QY 230 GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
DB 833 GGAAACTAATGAAGAGAAAGGTCAGCATTTTGAACCTTTATCATGAATTTTCATTAAT 892  
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
DB 893 TCTTGTGATGTTCCAAAGCTTAACCAAGTAGAGACAGACTGACCATCTTCAGCCCTACT 952  
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGlu 289  
DB 953 TCAGAAACCAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAA 1012  
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
DB 1013 AATACAGAGAGAAACAAATCAGAAATTTATTTATGGAGGTGATGATCTAGAGAGATC 1072

## RESULT 3

US-09-620-312D-1021  
Sequence 1021, Application US/09620312D  
Patent No. 6569662  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Ren, Feiyan  
APPLICANT: Chen, Rui-hong  
APPLICANT: Wehrman, Tom  
APPLICANT: Zhao, Qing A.  
APPLICANT: Xue, Aidong J.  
APPLICANT: Yang, Yonghong  
APPLICANT: Wang, Jian-Rui  
APPLICANT: Zhou, Ping  
APPLICANT: Ma, Yunqing  
APPLICANT: Wang, Dunrui  
APPLICANT: Wang, Zhiwei  
APPLICANT: John Tillinghast  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
TITLE OF INVENTION: Polypeptides  
FILE REFERENCE: 784CIP28  
CURRENT APPLICATION NUMBER: US/09/620,312D  
CURRENT FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1105  
SOFTWARE: pt\_fl\_genes Version 1.0  
SEQ ID NO 1021  
LENGTH: 1497  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (79)..(762)

QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrPheHis 329  
DB 1073 TATGATTACCTGATGATGTA----- 1093  
QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
DB 1093 ----- 1093  
QY 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369  
DB 1093 ----- 1093  
QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
DB 1094 -----GATGCTATATCTCTGAAACACACTGAACTCGCTCTCTCCAGATAAGCAAAA 1147  
QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
DB 1148 GGAGCAAAACAGACTTTTGAAGAAATGATGAATACATTCAGATCTCTTAGTCAAGTGT 1207  
QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnPro 429  
DB 1208 ATGTGTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTGATGGCTTACAGCAACA 1267  
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
DB 1268 GTACTCAACAGACACTGACTTATGTTTATTTGACATTTGATACAGAACTGTTTCCA 1327  
QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
DB 1328 GAGCTCAATAAGGTACAAAGAGGATTAACCTCTGTGACATCTTGATG 1375

## US-09-620-312D-1021

Alignment Scores:  
Pred. No.: 1.71e-06 Length: 1497  
Score: 136.50 Matches: 63  
Percent Similarity: 39.63% Conservative: 44  
Best Local Similarity: 23.33% Mismatches: 108  
Query Match: 5.60% Indels: 55  
DB: 10 Gaps: 10

US-09-744-313A-1 (1-465) x US-09-620-312D-1021 (1-1497)

QY 68 ProValGluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSer 87  
DB 25 CCGGTGAATGGGAAGATAGACCTACTACCTACTACTACTG----- 66  
QY 88 IleProTyrValaspPhePheGluaspProSerSerGluArgLysGluLysLysGluArg 107  
DB 67 -----GTTATGAAGTGAAGGAAAGAGCTAA 96  
QY 108 IleProValPheCysIleaspValGluArgAsnAspArgAlaValGlyHisGluPro 127  
DB 97 TTTACTGTATATAAATACTAGTAAGAAACCCAGAA----- 135  
QY 128 GluHisTrpSerValtyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThr 147  
DB 136 GAAAGCTGGGTATTTTCAGAGATACACTGACTTCTCTAGCTTAATGACAAATTAATA 195  
QY 148 GluPheHisGlyAlaPheProaspAlaGlnLeuProSerLysArgIleGlyProLys 167  
DB 196 GAGATGTTCCAGTTTCGATACACTCTCTCCAAACCGCTGTTAAAGATAATAC 255  
QY 168 AsnTyrGluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGluLysLeuLeu 187  
DB 256 AATGCTGACTTTTGAAGACAGACAAATTAGGATTACAAGCGTTCTTCAAAATTTAGTA 315  
QY 188 GlnHisProGluLeuSerAsnSerGlnLeuLeuAlaaspPheLeu-----SerPro 204  
DB 316 GCTCAGAGGACATTGCTAACTGCTGAGTGAAGAAATTTCTTTGTTGGATGATCCA 375  
QY 205 AsnGly-----GlyGluThrGlnPheLeuAspLysIleLeuProaspVal 219  
DB 376 CCGGTGCCATTGATAGCTAGAGGAAGACGAGGCACTCTGTGAACCTTTAGAGAGACA 435  
QY 220 AsnLeu-----GlyLysIleIleLysSerValProgly 230  
DB 435 AACTACCGCTTACAGAAAGAACTACTGTGAACAAAGAGAGATGGAATCACTAAAG-- 492  
QY 231 LysLeuMetLysGlyGlyGlnHisLeuGluPheIleMetAsnPheIleAsnSer 250  
DB 493 AACTGCTCAGTGAGAAGCACTTCATATAGACACT--TTAGAGACAGAAATCAGAAC 549  
QY 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
DB 550 TTG---TCTTTAGAACCTGAAGAATCA-----CTGATGTCTCAGAACAGAGGT 597  
QY 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290  
DB 598 GAACAGATCTTAAGGTGGAGCTCTCTGCACCTTGAGTTGATCAAGATGCTCTGATGAA 657  
QY 291 ThrGluArgLysGlnAsn-----GlnAsnTyrPheMet 301  
DB 658 GAATCTAGAGCTGATATAAACCATGCTTAAGTTTGTAGTGAACCTGAAATGCTGTATCA 717  
QY 302 GluValMetThrValGluGlyValTyrAsp 311  
DB 718 GAGATAGAAGTAGCAGAGAGTGGCATATGAT 747

## RESULT 4

US-08-625-322-1  
; Sequence 1, Application US/08625322  
; Patent No. 5804412  
; GENERAL INFORMATION:

APPLICANT: Gill, Gordon N.  
APPLICANT: Kurten, Richard C.  
APPLICANT: Cadena, Deborah L.  
TITLE OF INVENTION: Sorting Nexins and Methods of Using Same  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores  
STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/625,322  
FILING DATE: 01-APR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-UD 1955  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1974 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 45..1612  
US-08-625-322-1

## Alignment Scores:

Pred. No.: 2.96e-05 Length: 1974  
Score: 128.00 Matches: 98  
Percent Similarity: 34.79% Conservative: 69  
Best Local Similarity: 20.42% Mismatches: 187  
Query Match: 5.25% Indels: 126  
DB: 19 Gaps: 19

US-09-744-313A-1 (1-465) x US-08-625-322-1 (1-1974)

QY 26 GlySerLysIleLysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuPro--- 44  
DB 190 GGGCGCGCGTGGTCAATAAATCATCAGTCTCCAAAGATAACTACATCCCTTCTTCCATC 249  
QY 45 -----AsnTyrGlyValAlaGluGlyGluasp-----AspPhe 55  
DB 250 AACAAATGGTCCAAAGAAATGGATGATGATGATGATGATGATGATGATGATGATGATG 309  
QY 56 IleGluGluGlyIleValValMetGluAspSerProValGluAlaValSerThrPro 75  
DB 310 TTTGAGATGCCAGTGGAGCTATCTTGGACAGC-----ACACAAAT 354  
QY 76 AsnThrProArgAsnLeuAlaAlaTrpLysIleSerIleProTyrValAspPheGlu 95  
DB 355 AATCAGAAGAAGTGTAGCCAAACACTCATTTTCTCTCTCCTCAGGAAGCCACAAAT 414  
QY 96 -----AspProSerSerGluArgLysGluLysGluArgIleProValPhe 111  
DB 415 TCTTCGAAGCCCGAGCCACCTATGAGGAGCTAGAGGAAAGAAACAGAGGATCAATTT 474  
QY 112 CysIleaspValGluArgAsnAspArgAlaValGly----- 124  
DB 475 GATTTGACGTCGGTATACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 534

## RESULT 5

```

US-09-080-897-1
; Sequence 1, Application US/09080897
; Patent No. 5985574
; GENERAL INFORMATION:
; APPLICANT: King, Mary-Claire
; APPLICANT: Lynch, Eric D.
; APPLICANT: Lee, Ming
; APPLICANT: Morrow, Jan E.
; APPLICANT: Welch, Piri L.
; APPLICANT: Leon, Pedro E.
; TITLE OF INVENTION: Modulators of Actin
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESS: SCIENCE & TECHNOLOGY LAW GROUP
; STREET: 75 DENISE DRIVE
; CITY: HILLSBOROUGH
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94010
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/080,897
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: OSMAN, RICHARD A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UW97-001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 343-4341
; TELEFAX: (650) 343-4342
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3747 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3744
;
US-09-080-897-1
Alignment Scores:
Pred. No.: 0.00252 Length: 3747
Score: 116.00 Matches: 79
Percent Similarity: 40.10% Conservative: 75
Best Local Similarity: 20.57% Mismatches: 134
Query Match: 4.76% Indels: 96
DB: 2 Gaps: 20

US-09-744-313A-1 (1-465) x US-09-080-897-1 (1-3747)
Qy 98 SerSerGluArgLysGluLysLysGluArgGlieProValPheCysIleaspValGluArg 117
Db 2395 TCTGCCACGACCAAGACCAAGAGGATCAA----- 2424
Qy 118 AsnAspArgAlaValAlaGlyHisGluProGluHisTyrSerValTyrArgAspTyrLeu 137
Db 2425 -----GAAGTGTGAGAGAGAAAGAAATCTGTGCAAAAGAAAAAGTAAAA 2469
Qy 138 GluPheTyrValLeuGluSerLys-----LeuThrGluPheHisGlyAlaPhe 153
Db 2470 GAGTTAAAGGTGTGGATTCAAAGACGACCCAGACTCTCTCAATCTTTTGGGTCTCTTC 2529
Qy 154 ProAspAlaGlnLeuProSerLysArgGlieGlyProLysAsnTyrGluPheLeuLys 173
Db 2530 -----CGCATGCCCTATCAAGAGATT-----AAGAAATGTCATCTCGAGGTG 2571

```





2632 GAGCAGTAAAGATCTTCTGAAGTGAAGAGTAA-----TAT 2670  
213 AspLysLeuProAspValAsnLeuGlyLysLeuValProGly----- 230  
2671 GATGACCTGGTGAAGTCAAGAGATTTGGCGGGGATGGGACATGTCGCCGAGTGGG 2730  
231 -----LysLeu---MetLysGluLysGlyLysLeuGlu 241  
2731 CTTGGCTCAATGCCATTCTCTCAAGCTACAATTCAAGGAGGAGTGAAGAGTGAAGTTC 2790  
242 ProPheLeuMetAsnPheLeuSerCysGluSerProLysProLysProSerArgPro 261  
2791 CAGAGATTTGTTCTCTACTGCTGATGAGGAGTGAAGAGTGAAGTTC 2850  
262 -----GluLeuThrLeu-----SerProThrSerGluAsn 273  
2851 ATCTCTAGATTAACCTTGTGTTGGAATTAATCATGATGCTGGCTCCGAAATGCT 2910  
274 LysLysLeu---PheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsnThrGlu 292  
2911 GGTGCTTTGGCTTCAATATCAGTTCCTCTGTAAGTTCGAGACACCAAGTCCACAGAT 2970  
293 ArgLysGluAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyr 312  
2971 CAGAGATGAGTGTGTTACACTTCTGCTGATGAGTGTGAGATGAGTATCCCATGTC 3030  
313 LeuMetTyrValGlyArgValPheGlnValProAspTyrPheHisLeuLeuMet 332  
3031 CTCAGTTT-----CCAGACGAGCTTGGCCATGTGGAGAA 3066  
333 GlyThrArgLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCys 352  
3067 GCACGCGAGTCTCTCTGAAGC-----TTGCAAG 3099  
353 LysLeuGluGlnPheGlnGluHisArgLeuValSerLeuLeuThrLeuLeuArgAsp 372  
3100 AACCTAGATCAGATGAAGAACAA-----ATTCTGATGTGAAGTGTAT 3144  
373 AlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLys 392  
3145 GTT-----CAGAAATTTCCAGCTGCCAGATGAAGAAAGACAAAG----- 3183  
393 GlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGlu 412  
3184 ---TTGTGTAAGAAATGACCAAGCTTGTGAAGAT-----GCA 3219  
413 GluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsn 432  
3220 CAGAACAGTATTAACAGCTGGGATGATGATCTTAACATGAG-----ACCTCTAT 3273  
433 LysGlnLeu---ThrTyrValLeuLeuAsp-----IleValIleGlnLeuPhe 448  
3274 AAGGAGCTGGCGAGTACTTCTCTTTGACCCCAAGAGTGTCTGTGAAGAAATTTTC 3333  
449 ProGluLeuAsn 452  
3334 ATGGATCTTCAC 3345

STATE: D.C.  
COUNTRY: USA  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/899,595  
FILING DATE: 24-JUL-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-242701  
FILING DATE: 26-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 9-90170  
FILING DATE: 25-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Stephen A. Bent  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 049441/0112  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5822 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ORIGINAL SOURCE:  
ORGANISM: Human  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 28..3972  
US-08-899-595-4

Alignment Scores:  
Pred. NO.: 0.0106 Length: 5822  
Score: 113.50 Matches: 73  
Percent Similarity: 41.09% Conservative: 70  
Best Local Similarity: 20.98% Mismatches: 124  
Query Match: 4.66% Indels: 81  
DB: 3 Gaps: 19

US-09-744-313A-1 (1-465) x US-08-899-595-4 (1-5822)

Qy 134 ArgArgTyrLeuGluPheTyrValLeuGluSerLys-----LeuThrGluPhe 149  
Db 2686 AAAAAGTAAAGAGTAAAGGTTTGAAGTTCAGAGACAGCCAGAACTCTCAATCTTT 2745  
Qy 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
Db 2746 TTGGTTCCCTTC-----CGATGCCCTATCAAGAGATT-----AAGATGTC 2787  
Qy 170 GluPheLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeu---Gln 188  
Db 2788 ATCTCGAGGTGAATGAGGCTGTTCTGACTGAGTCTATGATCAGAACCTCATTAAGCAA 2847  
Qy 189 HisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGlu 208  
Db 2848 ATGCCAGAGCCAGAGCAGTAAAGATGCTTCTCAACTGAAGATGAA----- 2895  
Qy 209 ThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerVal 228  
Db 2896 -----TATGATGACCTGCTGAGTTCAGAGCAGTTTGGCTGGTGGTGGCAGCTGTG 2946  
Qy 229 ProGly-----LysLeu---MetLysGluLysGly 237  
Db 2947 CCCCAGTGGGGCTCGCCTCAATGCCATTTCTTCAAGCTACATTCAGCGAGCAAGTG 3006

RESULT 7  
US-08-899-595-4  
Sequence #4, Application US/08899595  
Patent No. 6111072  
GENERAL INFORMATION:  
APPLICANT: Narumiya, Shuh  
APPLICANT: Takahashi, No. 6111072uaki  
TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE  
TITLE OF INVENTION: ENCODING SAME  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W., Suite 500  
CITY: Washington

QY 238 GlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLys 257  
Db 3007 GAGATATCAGCCAGAGATGTTGCTGCTCAGCTGCAATGAGGATACAGAGT 3066  
QY 258 ProSerArgPro-----GluLeuThrIleLeu-----SerProThr 269  
Db 3067 GAGAGCTTTTCCATCTCCTAGAGATTACCTGCTTGTGAAATACATGAATGCTGGC 3126  
QY 270 SerGluAsnAsnLysLysLeu---PheAsnAspLeuPheLysAsnAsnAlaAsnArgAla 288  
Db 3127 TCCAGAAATGCTGGTCTTTGGCTTCAATATCAGCTTCTCTGTAAGCTTCGAGACACC 3186  
QY 289 GluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGly 308  
Db 3187 AAGTCCACAGATCAGAGATCAGCTGTTACCTTCTGCTGAGTGTGTGAGATGAC 3246  
QY 309 ValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLysHis 328  
Db 3247 TATCCGATGCTCCTCAAGTTT-----CCAGACGAGCTTGGC 3282  
QY 329 HisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyr 348  
Db 3283 CATGTGGAGAAAGCCAGCGAGTTTCTGCTGAAAC----- 3318  
QY 349 TyrLeuGlnCysLysLeuGlnLeuPheGlnGluHisArgLeuValSerLeuIleThr 368  
Db 3319 ---TTGCAAAAGAACCTAGATCAGATGAAGAACAA-----ATTCTGAT 3360  
QY 369 LeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGln 388  
Db 3361 GTGGAGCTGATGTT-----CAGAAATTTCCAGCTGCCACAGATGAAGAAC 3408  
QY 389 LysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuValLys 408  
Db 3409 AAG-----TTGTGTAAGAAATGACCAAGCTTTGTGAAGAT----- 3444  
QY 409 CysIleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGln 428  
Db 3445 ---GCACAGGACAGATTAACAAGCTGCGGATGATGCAATCTTAACATGGAG--- 3492  
QY 429 ProValLeuAsnLysGlnLeu---ThrTyrValLeuLeuAsp-----IleValIle 444  
Db 3493 ---ACCTCTATAGAGCTGGCGAGTACTTCTCTTTGACCCCAAGAAATGTCTGTT 3549  
QY 445 GlnGluLeuPheProGluLeuAsn 452  
Db 3550 GAAGAAATTTTCATGGAATCTTAC 3573

## RESULT 8

US-08-899-595-5/c

Sequence 5, Application US/08899595

Patent No. 611072

GENERAL INFORMATION:

APPLICANT: Narumiya, Shuh

APPLICANT: Takamashi, No. 611072uaki

TITLE OF INVENTION: RHO TARGET PROTEIN HUMAN MDIA AND GENE

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESS: Foley &amp; Lardner

STREET: 3000 K Street, N.W., Suite 500

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/899,595

FILING DATE: 24-JUL-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 8-242701  
FILING DATE: 26-AUG-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 9-90170  
FILING DATE: 25-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Stephen A. Bent  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 049441/0112  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5822 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
ANTI-SENSE: YES  
ORIGINAL SOURCE:  
ORGANISM: Human  
US-08-899-595-5

## Alignment Scores:

Pred. No.:	0.0106	Length:	5822
Score:	113.50	Matches:	73
Percent Similarity:	41.09%	Conservative:	70
Best Local Similarity:	20.98%	Mismatches:	124
Query Match:	4.66%	Indels:	81
DB:	3	Gaps:	19

US-09-744-313a-1 (1-465) x US-08-899-595-5 (1-5822)

QY 134 ArgArgTyrLeuGluPheTyrValLeuGluSerLys-----LeuThrGluPhe 149  
Db 3137 AAAAAGTAAAGAGTTTAAAGTGTGGATTCAAGACAGCCAGAAATCTCTCAATCTTT 3078  
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleGlyProLysAsnTyr 169  
Db 3077 TTGGGTCTCTTC-----CGCATGCCCTATCAGAGATT-----AAGATGTC 3036  
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnTyrLeuGlnLysLeuLeu---Gln 188  
Db 3035 ATCTGGAGGTGAATGAGCTGTTCTGACTGAGTCTATGATCCAGAACCTCATTAAAGCAA 2976  
QY 189 HisProGluLeuSerAsnSerGlnLeuLeuAlaPhePheLeuSerProAsnGlyGlyGlu 208  
Db 2975 ATGCCAGAGCCAGAGCAGTAAATAATGCTTTCTGAACCTGAAGAGTAA----- 2928  
QY 209 ThrGlnPheLeuAspLysIleLeuProAspValLeuLeuGlyLysIleLeuSerVal 228  
Db 2927 -----TATGATGACCTGCTGCTGAGTCAGAGCAGTTTGGCTGCTGATGGCACTGTG 2877  
QY 229 ProGly-----LysLeu---MetLysGluLysGly 237  
Db 2876 CCCCAGCTCGGCCTCGCTCAATGCCATTCTCTCAAGCTACAATTCCAGCAGCAATG 2817  
QY 238 GlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLys 257  
Db 2816 GAGATATCAGCCAGAGATGTTGTCTGCTCACTGCTGATGTGAGAGTTACGTAAGAGT 2757  
QY 258 ProSerArgPro-----GluLeuThrIleLeu-----SerProThr 269  
Db 2756 GAGAGCTTTTCCAAATCTCCTAGAGATTACCTTGTGTTGGAATATCATGAATCTCGC 2697  
QY 270 SerGluAsnAsnLysLysLeu---PheAsnAspLeuPheLysAsnAsnAlaAsnArgAla 288  
Db 2696 TCCAGAAATGCTGGTGTGCTTTTGGCTTCAATATCAGCTTCTCTGTGAAGCTTCGAGACACC 2637



Db 412393 AAGGTTAAA-----GAACCTGAACAAATTTGAT----- 412367  
QY 135 ArgTyrLeuGluPheTyrValLeuGluSer-----LysLeuThrGluPheHisGly 151  
Db 412366 -----GATTTCTATTCATTTGAAATTAACCAAGCTATCAATCCAGTTCATATA 412316  
QY 152 AlaPhe-----ProAspAlaGlnLeuProSerLysArgIleLeuGlyPro 166  
Db 412315 ACGATCCAATATGATCAGATGATCAACCTTTTGTGTTTAAAGGATCTTA----- 412262  
QY 167 LysAsnTyrGluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLys--- 185  
Db 412261 ---AAAGAGCAACATCCCAACAAAAAGTTGATGAATTTGGATGACTATATAATAACAAGAG 412205  
QY 186 ---LeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerPro 204  
Db 412204 CTTTATTCGAAATATGCTGATTTTAAACAA-----ATCGATGATCTTAAAGAG 412154  
QY 205 AsnGlyGluGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIle 224  
Db 412153 AATAATATGATCAAAATATTT-----GATTAGACAGGAA 412118  
QY 225 IleLysSerValProGlyLysLeuMetLysGlyLysGlyGlnHisLeu----- 240  
Db 412117 ATCGATGATCTTAAAGAGGAGTTATCTGAGGAAAAAGTAAGCACTTACATACCAAGAA 412058  
QY 241 -----GluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysPro 258  
Db 412057 TTACAGGATGATCTTTTTCAGAAAAATAGGGATTTATATGACCAACTCAATAATAACCA 411998  
QY 259 SerArgProGluLeuThrIleLeuSerProThrSer---GluAsnAsnLysLysLeuPhe 277  
Db 411997 -----GTTGCTATTATCTTTAACTGATGAAGTTAATGAAGAGTTGGAG 411953  
QY 278 Asn-----AspLeuPheLysAsnAsnAla 285  
Db 411952 AATTTAAACAAAGAAAAGCTTTTCTTCAGATCAATAGATGCTGAAAAATTAAGAGC 411893  
QY 286 AsnArgAlaGlu-----AsnThrGluArgLysGlnAsn 296  
Db 411892 AGTAATGTTTCAACAACTATAGCTTTTATACAGTTTAAACCAACCAATAACCACTT 411833  
QY 297 GlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrVal 316  
Db 411832 CAACACCACTTACTACACAGCAAGAGAACCAACCAAGATTTGGAT----- 411788  
QY 317 GlyArgValValPheGlnValProAspTyrIlePheHisLeuLeuMetGlyThrArgIle 336  
Db 411787 -----CTAGTTGAACAGAAAAATGATTTTGAAGAAATGACTG-----AAAAA 411743  
QY 337 LeuPheLysAsnThrLeu-----GluMetTyrThrAspTyrTyrLeuGlnCys 352  
Db 411742 CTTCCATGCAACACATCTAATGATGAGAAATGATAATGATGATCTTTTAAATCAGTAT 411683  
QY 353 LysLeuGluGlnLeuPheGlnGluHis----- 361  
Db 411682 GAACTA-----CTTTTGTATGAGAAATGATAATGATAATGATAATGATAATGATAAT 411629  
QY 362 -----ArgLeuValSerLeuIleThrLeuLeuArgAspAla 373  
Db 411628 CAAAGCCTAACTTAGATTTATCAAAAAAATTCAGCACTTAACATGAAATGATGTT 411569  
QY 374 IlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGln 393  
Db 411568 TTATTAGATGAATTCAGTGAACAGGCTTAAGATAATGACTTCAATAACACTAAAAAT 411509  
QY 394 ThrPheGluGluMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGlu 413  
Db 411508 TCATTGTGAAGACAGAAAAAGCACTCGATGAAACACT----- 411470  
QY 414 ThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnProValLeuAsnLys 433  
Db 411469 -----AATCGTTTAA-----ACAATTCAAAATCAA 411446

QY 359 GlnGluHisArgLeuValSerLeuIleThrLeuLeuArg-----AspAlaIle 374  
Db 259974 ACCAAACAGAGTATGATTAAGAAATCAGCACTTAAGAGTAAGATAGATGCTGAT 260033  
QY 375 PheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThr 394  
Db 260034 TTTTACATTAAAGCCGTTAGCTGAACCTTGATGATGATGATGATGATGATGATGAT 260087  
QY 395 PheGluGluMetMet-----AsnTyrIleProAspLeuLeuValLysCys 409  
Db 260088 TTTCAAGATGATTAATCAACAGCAAAATGCTCAACATGTTGAGGATAAGTTAGTTGCTTTA 260147  
QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
Db 260148 AACAAAGAAAGACCGTTTAAATACCAAAAGAGGCGCTTTTAACTTAAGACAATCT 260207  
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
Db 260208 -----GCTTTAATGATATC----- 260222  
QY 450 GluLeuAsnLysValGlnLysGlu 457  
Db 260223 -----AATAAACTCCAGCAGGAA 260240

## RESULT 10

US-08-545-528D-1/c

; Sequence 1; Application US/08545528D

; Patent No. 6537773

; GENERAL INFORMATION:

; APPLICANT: Fraser et al.

; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitium Genome, Fragment

; FILE OF INVENTION: Thereof, and Uses Thereof

; CURRENT APPLICATION NUMBER: US/08/545,528D

; CURRENT FILING DATE: 1995-10-19

; PRIOR APPLICATION NUMBER: US 08/488,018

; PRIOR FILING DATE: 1995-06-07

; PRIOR APPLICATION NUMBER: US 08/473,545

; PRIOR FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 1

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO: 1

; LENGTH: 580073

; TYPE: DNA

; ORGANISM: Mycoplasma genitalium

US-08-545-528D-1

Alignment Scores:

Pred. No.:	70.1	Length:	580073
Score:	109.50	Matches:	93
Percent Similarity:	34.68%	Conservative:	70
Best Local Similarity:	19.79%	Mismatches:	162
Query Match:	4.49%	Indels:	145
DB:	4	Gaps:	24

US-09-744-313A-1 (1-465) x US-08-545-528D-1 (1-580073)

QY 56 IleGluGluGlyIleValValMetGluAspSerProValGluAlaValSerThrPro 75  
Db 412543 CTTAAACCACTAGTACTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 412490  
QY 76 AsnThrProArgAsnLeuAlaValLysIleSerIleProTyrValAspPhePheGlu 95  
Db 412489 GCTACTAACAA-----GATCTTTTGAAGATTTGTTTAAAC 412454  
QY 96 AspProSerSerGluArgIleGlyLeu-----LysLysGluArgIleProValPheCysIleAsp 114  
Db 412453 CAACCTAGTAGAGAGATTAAATGAACAAAAAACCAAGAGTTCAAAATATTTCAACTGAT 412394  
QY 115 ValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSerValTyrArg 134  
Db -----

```

Qy 63 -----MetGluAspAspSerProValGluIaValSerThrProAsnThrProArgAsn 80
      ::::: |||::: :::: |||
Db 2844 CATAATGTTTCAGTTTGATGCTAATCTCTTGGCG-----GAAAT 2882

```

QY	81	Leu-----Ala	Ala	Trp	Lys	Lys	Leu	Ser	Leu	Pro	Tyr	Val	Asp	-----	92															
DB	2883	TTATTTT	GAAGGCTAT	GAGCTAT	GAGAAACCCCT	CGTGTG	TGATAC	CGGTG	CGAAT	TGGCC	CAG	2942																		
QY	93	---Phe	Phe	Glu	Asp	Pro	Ser	Ser	Glu	Arg	Lys	Glu	Lys	Val	Glu	Arg	Leu	Pro	Val	Phe	111									
DB	2943	GTCTTTT	TC-----	CCT	GAACT	GGA	-----	AA	TAT	AGCTT	GCC	GA	TTC	2984																
QY	112	Cys	Ile	Asp	Val	Glu	Arg	Asn	Asp	Arg	Ala	Val	Gly	His	Glu	Pro	Glu	His	Trp	Ser	131									
DB	2985	TGT-----	-----	CGA	GA	ATT	AGG	AAT	TTC	CT	CT	TAA	CAC	-----	3014															
QY	132	Val	Tyr	Arg	Arg	Tyr	Leu	Glu	Phe	Tyr	Val	Leu	Glu	Ser	Lys	Leu	Thr	Glu	Phe	His	151									
DB	3015	-----	-----	GC	A	C	A	C	A	C	A	-----	GC	A	C	A	C	A	C	A	3023									
QY	152	Ala	Phe	Pro	Asp	Ala	Gln	Leu	Pro	Ser	Lys	Arg	Ile	Leu	Gly	Pro	Lys	Asn	Tyr	Glu	171									
DB	3024	GCCTTT	CAGAT	CCCA	AGCTA	CAG	CAG	AA	TACT	CTCT	-----	TTT	3065																	
QY	172	Leu	Lys	Ser	Lys	Arg	Glu	-----	-----	-----	-----	-----	178																	
DB	3066	TTACG	GA	AA	AG	ATG	AC	CG	AGCT	CT	CT	CT	TG	GA	AG	CGCT	TG	TG	GA	AA	TTG	3125								
QY	179	-----	Phe	Gln	Lys	Tyr	Leu	-----	-----	-----	-----	-----	Gln	Lys	Leu	186														
DB	3126	GAGCGT	CTCT	AT	GAT	GAGT	CT	AC	TCT	AC	TCT	GGT	TAT	TG	AG	AAA	CTT	AT	CG	CA	CA	TTAT	3185							
QY	187	Leu	Gln	His	Pro	Glu	Leu	Ser	Asn	Ser	Gln	-----	-----	196																
DB	3186	CTGAGT	CTTCC	AG	ACT	TGGT	CCA	AGT	CT	CA	AGG	CT	TAT	TTA	AG	AAA	CGG	AA	GCT	TCT	3245									
QY	196	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	196																	
DB	3246	CTGG	ACC	CA	GA	AA	CT	AT	CT	CA	AG	ACT	TTT	CT	AAA	AT	TTT	CT	CT	GT	TG	AA	TTG	3305						
QY	197	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	Leu	Leu	Ala	Asp	-----	200												
DB	3306	GTG	AG	G	AG	G	AA	CA	GA	AA	AGT	TTT	TG	CT	AAA	AG	AGT	TGG	CT	TG	CT	TAT	TG	AA	AG	AA	CT	3365		
QY	201	-----	Phe	Leu	Ser	Pro	Asn	Gly	-----	Gly	Glu	Thr	Gln	Phe	Leu	Asp	Lys	Ile	Leu	216										
DB	3366	GTCT	CT	CT	GA	TTC	AA	CG	CG	CA	G	AG	G	AT	TG	G	AAA	ACC	-----	TAT	G	CT	AT	CT	CT	TTA	3419			
QY	217	Pro	Asp	Val	Asn	Leu	Gly	Lys	-----	Ile	Ile	Lys	Ser	Val	Pro	Gly	Lys	-----	231											
DB	3420	CCG	CT	TT	AT	CT	CA	AT	CC	AA	AG	CA	GC	CA	AA	T	TG	T	CT	T	AG	T	GT	TC	CG	CA	AA	AT	CT	3479
QY	232	-----	Leu	Met	Lys	Glu	Lys	Gly	Gln	His	Leu	Glu	Pro	Phe	Leu	Met	Asn	Phe	Ile	248										
DB	3480	CAAA	AT	CA	AA	T	CAT	CA	TG	AA	GA	GA	AG	CT	AA	AG	CT	CA	AG	GA	TG	T	CC	AT	CA	CA	GA	TAT	3539	
QY	249	Asn	Ser	Cys	Glu	Ser	Pro	Lys	Pro	Lys	Pro	Ser	Arg	Pro	Glu	Leu	Thr	Leu	Ser	Pro	258	</								

```

3765 GTGACGGAGATTTTGGAAACGTAGTCAAGAAAGGCGAGACTTGCAGACTT-----3818
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3767 LeuPheGlnGluHisArgLeuValSerLeuLeuThrLeuLeuArgAspAlaIlePheCys 376
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3819 TTAGTGAATATCAT-----GCCTATCTCTGATACCAAGACTTGAAGAT-----3860
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
377 GluAsnThrGluProArgSerLeuGlnAspGlnLysGlyAlaLysGlnThrPheGlu 396
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3861 -----AATCTGATTTGTCAGTGACCGT-----3884
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
397 GluMetMetAsnTyrIleProAspLeuValLysCysIleGlyGluGluThrLysTyr 416
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3885 -----TTACTGATT-----ATTGATGAAGTCCAAAG-----3911
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
417 GluSerIleArgLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThr 436
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3912 -----ATTTGTGTAGCTCTAGAAATCTGCTTCAAGAGACCTACGATATACAA-----3959
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
437 TyrValLeuLeuAspIleValIleGlnGluPheProGluLeuAsnLysValGlnLys 456
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
3960 ---TCTATTATCGATTAAATTGATAGGCTTTAGTAGGAGAGAAACAGGGTTCAACAA 4016
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
457 GluVal 458
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
4017 CGGATA 4022
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

RESULT 12
US-09-635-872A-21
; Sequence 21, Application US/09635872A
; Patent No. 6534300
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS FOR PRODUCING HIGHLY PHOSPHORYLATED LYSOSOMAL HYDROLASES
; FILE REFERENCE: 195613US0
; CURRENT APPLICATION NUMBER: US/09/635,872A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 21
; LENGTH: 3621
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-635-872A-21

Alignment Scores:
Pred. No.: 0.0225 Length: 3621
Score: 108.00 Matches: 104
Percent Similarity: 33.94% Conservative: 63
Best Local Similarity: 21.14% Mismatches: 153
Query Match: 4.43% Indels: 172
Dl: 4 Gaps: 29

US-09-744-313A-1 (1-465) x US-09-635-872A-21 (1-3621)

Qy 4 IleHisPheCysLeuIlePheArgAsnThrGlnLysArgGlyGlnSerPheGlyIleSer 23
Dl 1813 ATACATTATATCTACGCTTTCATAATACA-----AACGATGAAGATTC-----1857
Qy 24 ArgIleGlySerLysIleLysGlyValPheLysSerThrThrMetGluGlyAlaMetLeu 43
Dl 1858 -----AAATCCAGATACACAGTGGAGTGCACACAGGGAGGGACCAAACTG 1905
Qy 44 ProAsnTyrGlyValAlaGluGlyCysAspPheIleGlu-----57
Dl 1906 ---AATTCTAGCGCCAGAGGGTTACGAAAATTAGTTAGTCCCATACACACTTCTTCCA 1962
Qy 58 GluGlyIleValValMetCysAspSerProValGluAlaValSerThrProAsnThr 77
Dl 1963 GAGCGGGAATCTCTTTTGAGGAT-----ATTCCCAAGAA 1998
```

```

78 ProArgAsnLeuAlaIlePheLysIleSerIleProTyrValAspPhePheGluAspPro 97
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1999 AAACGC-----TTCCCGAAGTTTAAAGACACATGATTTAAC 2034
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
98 SerSerGluArgLysGlyLysGluArgIleProValPheCysIleAspValGluArg 117
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2035 TCAACAGAGAGAGCCAGGAGAGGTGAAATTCCTCTGGTAAATATTTTCACTCCTTCCA 2094
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
118 AsnAspArgAlaValGlyHisGluProGluHisTrpSerValTyrArgArgTyrLeu 137
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2095 AAAGACGCCAG-----TTG 2109
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
138 GluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGln 157
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2110 AGTCTCAATACCTTGATTCGAATCGAA-----CATGGA-----GACATCACT 2154
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
158 LeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGlu 177
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2155 TTG-----AAAGATACAATTTGTCCAAGTCA-----2181
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
178 GluPheGlnGluTyrLeuGlnLysLeuLeu-----GlnHisProGluLeuSerAsn 194
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2182 -----GCCTTGTCTGAGATCATTTCTGATGAACCTCACAGCATGCTAAATAAAAAAT 2232
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
195 SerGlnLeuLeuAlaAsp-----PheLeuSerProAsnGlyGlyGluThr 209
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2233 CAAGCTATAATAACAGATGAACAAATGACAGTTTGGTGGCTCCACAG-----GAAAA 2286
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLys-----226
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2287 CAGGTTTCAATAAAGCATCTTTCGCAAC-----AGCTTAGAGGTGCTGAAAGATTCACAGG 2343
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
227 ---SerValProGlyLysLeuMetLys-----GluLysGlyGlnHis-----239
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2344 TTGACTTTCTCTGCGATCAGTGAATAAGTGAATGCTATGACACAGGCTCAGAAATCCACC 2403
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
240 -----LeuGluProPheIleMetAsnPheIleAsn 249
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2404 CTGGACTTGGAGACACACAGCAAGATTAGAGTGAACACTCACCCAAAAAACCATAGGC 2463
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2464 GGA---AATGTGACAAAGAAAGCC-----CCATCTCTGATGTTTCCACTGGAAAGC 2514
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2515 CAGATGACAAAAGAAAGAAATACAGGGAAAGAAAGAGAACAGTAGAATGAGGAA 2574
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2575 AATGCTGAAAAT-----2586
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHis 329
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2587 -----CACATAGGC-----2595
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2596 -----GTTACTGAAGTGTACTTGGAGAAAGCTGCAGCATACACAGATAGTTAC 2646
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
350 Leu-----GlnCysLysLeuGluGlnLeuPheGlnGluHis 361
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2647 TTGGGCTTTTGGCATGGAGAAAAAAGTATTTCCTAGATCTTCTCGAGAAAGAGAG 2706
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
362 ArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluPro 381
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2707 TCATTGAAGACACAATTTGGCATCTTCACTGATAGC-----AAGAACTACT---GGG 2754
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
382 ArgSerLeuGlnAsnLysGlyAlaLysGlnThrPheGluGluMetAsnTyr 401
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
2755 AGGCACTAAAGAT-----ACATTTGCAGATTCCTCCAGATAT 2793
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
402 IleProAspLeuLeuValLysCysIleGlyGluThrLysTyrGluSerIleArgLeu 421
```



Db 2794 GTAAATAAATCTTAAATAGCAAGTTTGA 2823  
Qy 422 LeuPheaspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuasp 441  
Db 2824 ---TTACATCGCGAAGTCCCTGCT---CACATGCTCACATGATGACCGG 2871  
Qy 442 IleValIleGlnGluLeu---PheProGlu 450  
Db 2872 ATTGTATGCAAGACTGCAAGATATGTTCCCTGAA 2907

## RESULT 13

US-09-636-077A-21  
; Sequence 21, Application US/09636077A  
; Patent No. 653785  
; GENERAL INFORMATION:  
; APPLICANT: CANFIELD, WILLIAM  
; TITLE OF INVENTION: METHODS OF TREATING LYSOSOMAL STORAGE DISEASE  
; FILE REFERENCE: 195612US0  
; CURRENT APPLICATION NUMBER: US/09/636,077A  
; CURRENT FILING DATE: 2000-08-10  
; PRIOR APPLICATION NUMBER: 60/153,831  
; PRIOR FILING DATE: 1999-09-14  
; NUMBER OF SEQ ID NOS: 52  
; SOFTWARE: Patent in version 3.0  
; SEQ ID NO 21  
; LENGTH: 3621  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-636-077A-21

Alignment Scores:  
Pred. No.: 0.0225 Length: 3621  
Score: 108.00 Matches: 104  
Percent Similarity: 33.94% Conservative: 63  
Best Local Similarity: 21.14% Mismatches: 153  
Query Match: 4.43% Indels: 172  
DB: 4 Gaps: 29

US-09-744-313A-1 (1-465) x US-09-636-077A-21 (1-3621)

Qy 4 IleHisPheCysLeuIlePheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSer 23  
Db 1813 ATACATTTTAACTCAGCTTCAAAATACA-----AAGATGAAGTTC----- 1857  
Qy 24 ArgIleGlySerLysIleLysGlyValPheLysSerThrThrMetGluGlyAlaMetLeu 43  
Db 1858 -----AAATGCAGATAACAGTGGAGGTGGACACACAGGAGGACCAAACTG 1905  
Qy 44 ProAsnTyrGlyValAlaGluGlyGluAspAspPheIleGlu----- 57  
Db 1906 ---AATTCACGGCCACAGAGGTTACGAAATTTAGTTAGTCCATAACACTTCTTCCA 1962  
Qy 58 GluGlyIleValValMetGluAspAspSerProValGluAlaValSerThrProAsnThr 77  
Db 1963 GAGCGGAAATCCCTTTTGAGAT-----ATTCCCAAGAA 1998  
Qy 78 ProArgAsnLeuAlaIlePheLysIleSerIleProTyrValAspPheGluAspPro 97  
Db 1999 AAACGC-----TTCCGAAGTTTAAGACATGATGTTAAC 2034  
Qy 98 SerSerGluArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArg 117  
Db 2035 TCAACAGGAGAGCCAGGAGAGGTGAAATTCCTCGTAAATATTTTCACTCCCTTCCA 2094  
Qy 118 AsnAspArgAlaValAlaGlyHisGluProGluHisIlePheSerValTyrArgArgTyrLeu 137  
Db 2095 AAAGACCCCGAG-----TTG 2109  
Qy 138 GluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGln 157  
Db 2110 AGTCTCAATACCTTGGATTGGCAACTGGAA-----CATGGA-----GACATCACT 2154

Qy 158 LeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGlu 177  
Db 2155 TTG-----AAAGGATACAAATTTGTCCAGTCA----- 2181  
Qy 178 GluPheGlnGluTyrLeuGlnLysLeuLeu-----GlnHisProGluLeuSerAsn 194  
Db 2182 -----GCCTTGCTGAGATCATTTCTGATGAATCATCAGCATGCTAAATAAAAAAT 2232  
Qy 195 SerGlnLeuLeuAlaAsp-----PheLeuSerProAsnGlyGlyGluThr 209  
Db 2233 CACGCTATAATACAGATGAACAAATGACAGTTTGTGGCTCCACAG-----GAAAAA 2286  
Qy 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLys----- 226  
Db 2287 CAGGTTCAATAAAGCACTTGGCCAAAC---AGCTTAGAGTGTCTGAAAGATTGCAGAGG 2343  
Qy 227 ---SerValProGlyLysLeuMetLys-----GluLysGlyGlnHis----- 239  
Db 2344 TTGACTTTTCTCGCAGTGAGTGAATAAGTGAATGCTCATCAGCAGGCTCAGATCCACCC 2403  
Qy 240 -----LeuGluProPheIleMetAsnPheIleAsn 249  
Db 2404 CTGGACTTGGAGACCACAGCAAGATTTAGAGTGGAACTCACACCCAAAAAACCATAGGC 2463  
Qy 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
Db 2464 GGA---AATGTGACAAAGAAAGAGCCC-----CCATCTCTGATTGTTCCTCGGAAGC 2514  
Qy 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
Db 2515 CAGATGACAAAGAAAGAAATTCACAGGGAAGAAAGAGAAACAGACAGTAGAATGGAGAA 2574  
Qy 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
Db 2575 AATGCTGAAAT----- 2586  
Qy 310 TyrAspTyrLeuMetTyrValGlyArgValPheGlnValProAspTyrLeuHis 329  
Db 2587 -----CACATAGGC----- 2595  
Qy 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
Db 2596 -----GTTACTGAAGTGTACTTGGAAAGAAAGCTGCAGCATTACACAGATAGTAC 2646  
Qy 350 Leu-----GlnCysLysLeuGluGlnLeuPheGlnGluHis 361  
Db 2647 TTGGCTTTTTCATCGGAGAAAAAAGTATTTCTAGATCTTCTCGACGAAGAGAG 2706  
Qy 362 ArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluPro 381  
Db 2707 TCATTGAAGACACAAATTTGGCATACTTCACCTGATAGC-----AAGATACT---GGG 2754  
Qy 382 ArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyr 401  
Db 2755 AGGCACTTAAAGAT-----ACATTTGCAGATTCCTCCAGATAT 2793  
Qy 402 IleProAspLeuLeuValLysCysIleGlyGluThrLysTyrGluSerIleArgLeu 421  
Db 2794 GTAAATAAATCTTAAATAGCAAGTTTGA----- 2823  
Qy 422 LeuPheaspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAsp 441  
Db 2824 ---TTCATCGCGAAGATCCCTGCT---CACATGCTCATCATGATTGACCGG 2871  
Qy 442 IleValIleGlnGluLeu-----PheProGlu 450  
Db 2872 ATTGTATGCAAGAACTGCAAGATATGTTCCCTGAA 2907

## RESULT 14

US-09-635-872A-20  
; Sequence 20, Application US/09635872A  
; Patent No. 6534300  
; GENERAL INFORMATION:

```

RESULT 15
US-09-636-077A-20
; Sequence 20, Application US/09636077A
; Patent No. 6537785
; GENERAL INFORMATION:
; APPLICANT: CANFIELD, WILLIAM
; TITLE OF INVENTION: METHODS OF TREATING LYSSOMAL STORAGE DISEASE
; FILE REFERENCE: 195612US0
; CURRENT APPLICATION NUMBER: US/09/636,077A
; CURRENT FILING DATE: 2000-08-10
; PRIOR APPLICATION NUMBER: 60/153,831
; PRIOR FILING DATE: 1999-09-14
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 3783
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-636-077A-20
Alignment Scores:

```

```
Pred. No.: 0.0242 Length: 3783
Score: 108.00 Matches: 104
Percent Similarity: 33.94% Conservative: 63
Best Local Similarity: 21.14% Mismatches: 153
Query Match: 4.43% Indels: 172
DB: 4 Gaps: 29

US-09-744-313A-1 (1-465) x US-09-636-077A-20 (1-3783)

QY 4 IleHisPheCysLeuIlePheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSer 23
DB 1837 ATACATTTTAACTCTACGTTTCAAAATACA-----AACGATGAAGAGTTC-----1881
QY 24 ArgIleGlySerLysIleLysGlyValPheLysSerThrThrMetGluGlyAlaMetLeu 43
DB 1882 -----AAATGCGAGTAACACAGTGGAGGTGGACACAGGAGGAGGACAAACTG 1929
QY 44 ProAsnTyrGlyValAlaGluGlyGluAspAspPheIleGlu-----57
DB 1930 ---AATTCACGCGCCAGAGGGTTACGAAATTTAGTTAGTCCCATACACTTCTTCCA 1986
QY 58 GluGlyIleValValMetGluAspAspSerProValGluAlaValSerThrProAsnThr 77
DB 1987 GAGCGGAAATCCTTTTTCAGGAT-----ATTCCTCAAGAA 2022
QY 78 ProArgAsnLeuAlaAlaTrpLysIleSerIleProTyrValAspPheGluAspPro 97
DB 2023 AAACGC-----TTCCCGAAGTTTAAAGACATGATGTTAAC 2058
QY 98 SerSerGluArgLysGlyLysGluArgIleProValPheCysIleAspValGluArg 117
DB 2059 TCACAGAGGAGACCCAGGAGAGTGAATTCCTCGTAAATATTTCACTCTCTCCA 2118
QY 118 AsnAspArgAlaValGlyHisGluProGluHisTrpSerValTyrArgTyrLeu 137
DB 2119 AAAGAGCCCGAG-----TTG 2133
QY 138 GluPheTyrValLengLuserLysLeuThrGluPheHisGlyAlaPheProAspAlaGln 157
DB 2134 AGTCTCAATACCTTGGATTGCAACTGGAA-----CATGGA-----GACATCACT 2178
QY 158 LeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGlu 177
DB 2179 TTG-----AAAGATACAATTTGCTCCAGTCA-----2205
QY 178 GluPheGlnGluTyrLeuGlnLysLeuLeu-----GlnHisProGluLeuSerAsn 194
DB 2206 -----GCCTTGCTGAGATCATTTCTGATGAACCTCACAGATGCTAAATAAATAAT 2256
QY 195 SerGlnLeuLeuAlaAsp-----PheLeuSerProAsnGlyGlyGluThr 209
DB 2257 CAAGCTATAATAACAGATGAACAAACAAATGACAGTTTGGTGGCTCCACAG-----GAAAAA 2310
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLys-----226
DB 2311 CAGGTTCAATAAAGCATCTTGGCAAAAC---AGCTTAGAGTGTCTCAAAAGATTGACAGAG 2367
QY 227 ---SerValProGlyLysLeuMetLys-----GluLysGlyGlnHis-----239
DB 2368 TTGACTTTTCTGAGTGAAGTGAATAAGTGAATGGTCAAGCAGGCTCAGATCCACCC 2427
QY 240 -----LeuGluProPheIleMetAsnPheIleAsn 249
DB 2428 CTGGACTTGGAGACCACAGCAAGNTTTAGAGTGGAAACTCACACCCAAAAAACCATAGGC 2487
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269
DB 2488 GGA---AATGTGACAAAGAAAGAAAGCCC-----CCATCTCTGATTTGCTCCACTGGAAAGC 2538
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289
DB 2539 CAGATGCAAAAGAAAGAAAGAAATTCACAGGGAAGAAAGAAAGAAAGAAAGTGAATGGAGGAA 2598
```

```
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309
DB 2599 AATGCTGAAAT-----2610
QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHisHis 329
DB 2611 -----CACATAGGC-----2619
QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349
DB 2620 -----GTTACTGAGTGTACTTGGAGAAAGCTGACGATTACACAGTAGTTAC 2670
QY 350 Leu-----GlnCysLysLeuGluGlnLeuPheGlnGluHis 361
DB 2671 TTGGGCTTTTGGCATGGGAGAAAAAAGTATTTCTAGATCTTCTCGACGAGAAGAG 2730
QY 362 ArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluPro 381
DB 2731 TCATTGAAGACACAAATTGGCATACTTCACTGATAGC-----AAGATACT--GGG 2778
QY 382 ArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyr 401
DB 2779 AGCAACTAAAGAT-----ACATTCGAGATTCCTCAGATAT 2817
QY 402 IleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArgLeu 421
DB 2818 GTAATATAAATCTTAATAGCAAGTTTGG-----2847
QY 422 LeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAsp 441
DB 2848 ---TTCACATCGCGGAAAGTCCCTGCT-----CACATGCTCAGATGATTGACCGG 2895
QY 442 IleValIleGlnGluLeu-----PheProGlu 450
DB 2896 ATTGTTATGCAAGAACTGCAAGATATGTTCTCCTGAA 2931
```

Search completed: January 31, 2004, 05:18:33  
Job time : 523 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 06:31:47 ; Search time 86 Seconds  
(without alignments)  
2386.548 Million cell updates/sec

Title: US-09-744-313A-1  
Perfect score: 465  
Sequence: 1 MYLRLCLIFRNTQKRGESF.....ELFPELNKVKQKVTSTWMM 465

Scoring table: OLIGO

Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 569978 seqs, 220691566 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1135299

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DSV=xlh  
-Q=/cg2\_1/USPTO.spool/US09744313/runat\_30012004.113613.4617/app\_query.fasta\_1.647  
-DB=Issued Patents NA -QPMF=fastap -SUFFIX=rni -MINMATCH=0.1 -IOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODES=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313 -CGN\_1\_1\_56=brunat\_30012004.113613.4617 -NCFPU=6 -ICFPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=6 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:  
1: /cg2\_6/ptodata/2/ina/5A COMB.seq:  
2: /cg2\_6/ptodata/2/ina/5B COMB.seq:  
3: /cg2\_6/ptodata/2/ina/6A COMB.seq:  
4: /cg2\_6/ptodata/2/ina/6B COMB.seq:  
5: /cg2\_6/ptodata/2/ina/PTUS COMB.seq:  
6: /cg2\_6/ptodata/2/ina/backfiles1.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	455	98.1	1716	4	US-09-620-312D-922
2	307	66.0	1551	4	US-09-620-312D-923
3	8	1.7	85	2	US-08-332-766A-44
4	8	1.7	740	4	US-09-342-681C-99
5	8	1.7	966	4	US-09-252-991A-9495
6	8	1.7	1239	4	US-09-252-991A-9666
7	8	1.7	1842	4	US-09-205-258-160
8	8	1.7	3043	3	US-09-008-271A-14
9	8	1.7	152331	3	US-09-128-155-16
10	8	1.7	176373	3	US-09-128-155-17
11	7	1.5	161	1	US-08-110-786A-4
12	7	1.5	161	1	US-08-110-786A-6

c 13	7	1.5	223	1	US-08-435-684A-43	Sequence 43, Appl
c 14	7	1.5	223	2	US-08-934-877A-43	Sequence 43, Appl
c 15	7	1.5	223	3	US-08-871-678C-43	Sequence 43, Appl
c 16	7	1.5	276	4	US-09-252-991A-991	Sequence 991, Appl
c 17	7	1.5	280	4	US-09-313-294A-998	Sequence 998, Appl
c 18	7	1.5	285	4	US-09-107-532A-3549	Sequence 3549, Appl
c 19	7	1.5	288	4	US-09-313-294A-3533	Sequence 3533, Appl
c 20	7	1.5	288	4	US-09-313-294A-5146	Sequence 5146, Appl
c 21	7	1.5	321	1	US-08-175-388-2	Sequence 2, Appl
c 22	7	1.5	321	2	US-08-779-620-2	Sequence 2, Appl
c 23	7	1.5	321	2	US-08-818-726-2	Sequence 2, Appl
c 24	7	1.5	321	3	US-09-018-584A-20	Sequence 20, Appl
c 25	7	1.5	365	3	US-09-326-039-1	Sequence 1, Appl
c 26	7	1.5	370	2	US-08-332-766A-8	Sequence 8, Appl
c 27	7	1.5	371	3	US-09-326-039-11	Sequence 11, Appl
c 28	7	1.5	421	2	US-08-332-766A-25	Sequence 25, Appl
c 29	7	1.5	475	4	US-09-221-017B-561	Sequence 561, Appl
c 30	7	1.5	597	2	US-08-332-766A-19	Sequence 19, Appl
c 31	7	1.5	627	3	US-09-328-111-194	Sequence 194, Appl
c 32	7	1.5	657	4	US-09-252-991A-13015	Sequence 13015, A
c 33	7	1.5	661	2	US-08-529-878B-37	Sequence 37, Appl
c 34	7	1.5	666	4	US-09-252-991A-11270	Sequence 11270, A
c 35	7	1.5	702	4	US-09-252-991A-6791	Sequence 6791, Appl
c 36	7	1.5	710	3	US-08-998-416-982	Sequence 982, Appl
c 37	7	1.5	725	3	US-08-998-416-984	Sequence 984, Appl
c 38	7	1.5	765	4	US-09-252-991A-6948	Sequence 6948, Appl
c 39	7	1.5	780	4	US-09-252-991A-6722	Sequence 6722, Appl
c 40	7	1.5	784	4	US-09-205-258-101	Sequence 101, Appl
c 41	7	1.5	797	3	US-08-981-803-13	Sequence 13, Appl
c 42	7	1.5	797	3	US-08-981-803-27	Sequence 27, Appl
c 43	7	1.5	797	3	US-08-983-440-13	Sequence 13, Appl
c 44	7	1.5	797	3	US-08-983-440-27	Sequence 27, Appl
c 45	7	1.5	797	4	US-09-367-895-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1

US-09-620-312D-922  
; Sequence 922, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yungqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 656962el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes Version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716  
; TYPE: DNA

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71)..(1543)
US-09-620-312D-922

Alignment Scores:
Pred. No.: 0 Length: 1716
Score: 456.00 Matches: 456
Percent Similarity: 100.00% Conservative: 0
Best/Local Similarity: 100.00% Mismatches: 0
Query Match: 98.06% Indels: 0
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-620-312D-922 (1-1716)

QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29
Db 173 TTTTCGGAACACACAGAAAGGGGAGATCATTTTGGAAATCAGCAGAAATAGGTAGCAAAAT 232
QY 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnThrGlyValAla 49
Db 233 AAAGAGATTTCCAAAGATACCAATAGGAGGAGCTATGTTGCCCTAAATATAGGTGTAGCT 292
QY 50 GluGlyGluAspPheIleGluGlyIleValValMetGluAspSerProVal 69
Db 293 GAAGGTGAGATGATTTTATTTGAAGAAGGTATTTGTTGAATGGAAGATGATTTCCAGTG 352
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaIleLysIleSerIlePro 89
Db 353 GAGGCTGTGAGCACACCTAAATCTCCCGAAACCTTGCTGCATGCAAAATATAGCATTTCCA 412
QY 90 TyrValAspPheGluAspProSerSerGluArgLysGluLysLysGluArgIlePro 109
Db 413 TATGTAGACTTTTTTGGAGATCCCTCTCTGAAAGGAAGGAGAAAGAAAGAAATTCCT 472
QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129
Db 473 GGTGTTTGTATTCATGTTGAAGAATAATGATAGAGAGCAGGTGGACAGACCTTGACAT 532
QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149
Db 533 TGGTCTGCTCTAGAGATATCTTGAATCTATGTACTTGAATCAAACTAAACAGAAATTT 592
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169
Db 593 CATGTTGATTTCTCTGATGCCAGCTTCTCTTAAGAGGATCAATGGCCCCCAAAATAT 652
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGluHis 189
Db 653 GAATTTCTTAAAGTCAAAGGGAAGGATTCCAAGAAATATCTACAGAACTTTCTGCAGCAT 712
QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209
Db 713 CCAGAACTGAGTAATAGTCACTTCTGGCAGACTTTCTTCCCTAATGTTGGGGAACA 772
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro 229
Db 773 CAATTTCTGTAAGATACCTACAGATGTAAATCTTTGGAAAAATATATAAAATCTGTTCT 832
QY 230 GlyLysLeuMetLysGlyGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn 249
Db 833 GGAAAACTAATGAAGAGAAAGGTGAGCATTTTGGAACTTTTATCATGAAATTTTCAATA 892
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269
Db 893 TTTTGTGAGTCTCCAAAGCCTTAACCAAGTAGACAGCACTGACCATTTCTCAGCCCTACT 952
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289
Db 953 TCAGAAAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGGTCTGAA 1012
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309
```

```

Db 1013 AATACAGAGAGAAAGCAAAATCAGAAATATATTTATGGAGGTGATGACTGTAGAGAGTC 1072
QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHis 329
Db 1073 TATGATTTACCTGATGTATGTAGGAGGGTAGTTTCCAGGTTCCCTGACTGGCTTCATCAT 1132
QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349
Db 1133 CTCCTTAATGGGAATCGAATCTCTTTAAAAACACCCCTGGAATGTATGACTGATTACTAT 1192
QY 350 LeuGlnCysLeuLysGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369
Db 1193 CTTCACTGTAAACTAGAACAGCTATTTTCCAGGAGCACCGTTTGGTCTCACTCATAACACTT 1252
QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389
Db 1253 CTCAGAGATGCTATATTTCTGTGAAAACACCTGAACCTCGCTCTCTCAAGATTAAGCAAAA 1312
QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409
Db 1313 GGAGCAAAACACACTTTTGAAGAAATGATGAATTTACATCCAGATCTGTTAGTCAAGTGT 1372
QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429
Db 1373 ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGCTTACAGCAACA 1432
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449
Db 1433 GTACTCAACAACAGCAGCTGATTTATGTTTATTTGGACATTTGTGATACAGAACTGTTTCCA 1492
QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465
Db 1493 GAGCTCAATAGGTACAAAAGGAAGTTACCTCTGTGACATCTTTGGATG 1540

RESULT 2
US-09-620-312D-923
; Sequence 923, Application US/09620312D
; Patent No. 6569662
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyang
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunding
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: John Tillinghast
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and
; FILE REFERENCE: 784CIP28
; CURRENT APPLICATION NUMBER: US/09/620,312D
; CURRENT FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1105
; SOFTWARE: pt.PL_genes version 1.0
; SEQ ID NO 923
; LENGTH: 1551
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

NAME/KEY: CDS  
LOCATION: (71)..(1378)  
US-09-620-312D-923

Alignment Scores:

Pred. No.: 3 15e-306 Length: 1551  
Score: 307.00 Matches: 307  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 66.02% Indels: 0  
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-620-312D-923 (1-1551)

QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerIle 29  
Db 173 TTTCGGAACACACAGAAAAGGGAGAAATCATTTTGGAAATCAGCAGAAATAGTACCAAAATT 232  
QY 30 LysGlyValPheLysSerThrMetGluGlyAlaMetLeuProAsnThrGlyValala 49  
Db 233 AAAGGAGTATTCAAAAGTACCAATGGAGGGAGCTATGTCCTAAATATGCTAGCT 292  
QY 50 GluGlyGluAspPheIleGluGlyIleValMetGluAspSerProVal 69  
Db 293 GAAGGTGAGATGATTTTATGGAAGAGTATTTGTAATGGAAGATGATTCGAGTG 352  
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro 89  
Db 353 GAGGCTGTGAGCACACCTAATATCTCCCGAAACCTTGCTGCATGGAAATAGCATCCA 412  
QY 90 TyrValaspPheGluAspProSerSerGluArgLysGluLysGluArgIlePro 109  
Db 413 TATGTAGACTTTTTCGAGTCCCTCTCTGAAAGAGGAGGAGAGAGAGAGAGAGAGAGAG 472  
QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
Db 473 GTGTTTGTATTCATGTTGAAGAAATGATAGAGAGCAGTGGACAGAGCTGAACAT 532  
QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
Db 533 TGCTCTGTCTATGAGAGATTCCTGATTTCTATGTAATCTGATCAAACTAAGAAATTT 592  
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleGlyProLysAsnTyr 169  
Db 593 CATGTTGATTTCTCTGATGCCAGCTTCTCTTCAAGAGATCATTTGCCCCCAAAATAT 652  
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHis 189  
Db 653 GAATTTCTTAAGTCAAAGAGGAGAGATTCAGAGATATCTACAGAAATCTCTGCAGCAT 712  
QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
Db 713 CCAAGACTGAGTATATAGTCAACTTCTGGCAGACTTTCTTCCCTTAATGTTGGGAAACA 772  
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro 229  
Db 773 CAATTTCTGATAGATACTACAGATGTAAATCTTGGGAAATATATAAATCTGTCTCT 832  
QY 230 GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPhelLeuAsn 249  
Db 833 GGAATACTAATGAAGAGAGAGAGGTCAGCAATTTGGAACCTTTTATCATGATTTTCAAT 892  
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
Db 893 TCTTGTGAGTCTCAAGAGCCTAAACCAAGTAGACAGAACTTGACCACTTCTCGCCCTACT 952  
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
Db 953 TCAGAAACACACAGAGAGCTTTTCAATGATCTCTTTAAATAATCAACACCTGCTGAA 1012  
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
Db 1013 AATACAGAGAGAGAGCAATCATGATTTTATGAGGTGATGATCTAGAGAGATC 1072

QY 310 TyrAspTyrLeuMetTyrVal 316  
Db 1073 TATGATTACCTGATGATGTA 1093

RESULT 3

US-08-332-766A-44  
; Sequence 44, Application US/08332766A  
; Patent No. 5843647  
; GENERAL INFORMATION:  
; APPLICANT: JEFFREYS, Alec J.  
; APPLICANT: ARMOUR, John  
; TITLE OF INVENTION: SIMPLE TANDEM REPEATS  
; NUMBER OF SEQUENCES: 125  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN DARBY & CUSHMAN, L.L.P.  
; STREET: 1100 New York Avenue, N.W.  
; CITY: Washington  
; STATE: D. C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3918  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/332,766A  
; FILING DATE: 01-NOV-1994  
; CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9326052.9  
; FILING DATE: 21-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BIRD, Donald J. 25,323  
; REGISTRATION NUMBER: 25,323  
; REFERENCE/DOCKET NUMBER: 217211/M94/0434/GB  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 6714627 CUSH  
; INFORMATION FOR SEQ ID NO: 44:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 85 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-08-332-766A-44

Alignment Scores:  
Pred. No.: 6.59 Length: 85  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.72% Indels: 0  
DB: 2 Gaps: 0

US-09-744-313A-1 (1-465) x US-08-332-766A-44 (1-85)

QY 100 GluArgLysGluLysLysGluArg 107

Db 49 GAAAGAAAGAAAGAAAGG 72

RESULT 4

US-09-342-681C-99  
; Sequence 99, Application US/09342681C  
; Patent No. 6355782  
; GENERAL INFORMATION:  
; APPLICANT: Zonana et al.  
; TITLE OF INVENTION: Hypohydrotic ectodermal dysplasia genes and proteins  
; FILE REFERENCE: 52978  
; CURRENT APPLICATION NUMBER: US/09/342,681C



; CURRENT FILING DATE: 1999-06-29  
; PRIOR APPLICATION NUMBER: 60/092,279  
; PRIOR FILING DATE: 1998-07-09  
; PRIOR APPLICATION NUMBER: 60/112,366  
; PRIOR FILING DATE: 1998-12-15  
; NUMBER OF SEQ ID NOS: 123  
; SOFTWARE: Patent in Ver. 2.1  
; SEQ ID NO 99  
; LENGTH: 740  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-342-681C-99

Alignment Scores:  
Pred. No.: 50.3 Length: 740  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.72% Indels: 0  
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-342-681C-99 (1-740)

QY 235 GluysGlyGlnHisLeuGlnPro 242  
DB 515 GAAAGGGGCGACATCTGAGCCT 538

## RESULT 5

US-09-252-991A-9495/c  
; Sequence 9495, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 9495  
; LENGTH: 966  
; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9495

Alignment Scores:  
Pred. No.: 64.6 Length: 966  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.72% Indels: 0  
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-252-991A-9495 (1-966)

QY 422 LeuPheAspGlyLeuGlnPro 429  
DB 351 CTCTTCGACGGTTTCGACGACCA 328

## RESULT 6

US-09-252-991A-9666

; Sequence 9666, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 9666  
; LENGTH: 1239  
; TYPE: DNA  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-9666

Alignment Scores:  
Pred. No.: 81.5 Length: 1239  
Score: 8.00 Matches: 8  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 1.72% Indels: 0  
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-252-991A-9666 (1-1239)

QY 422 LeuPheAspGlyLeuGlnPro 429  
DB 661 CTCTTCGACGGTTTCGACGACCA 684

## RESULT 7

US-09-205-258-160  
; Sequence 160, Application US/09205258  
; Patent No. 6525174

; GENERAL INFORMATION:  
; APPLICANT: Young et al.

; TITLE OF INVENTION: 207 Human Secreted Proteins  
; FILE REFERENCE: P2007P1  
; CURRENT APPLICATION NUMBER: US/09/205,258

; CURRENT FILING DATE: 1998-12-04  
; EARLIER APPLICATION NUMBER: PCT/US98/11422

; EARLIER FILING DATE: 1998-06-04  
; EARLIER APPLICATION NUMBER: 60/048,885

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,375

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,881

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,880

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,896

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/049,020

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,876

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,895

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,884

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,894

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,971

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,964

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,882

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,899

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,893

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,900

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,901

; EARLIER FILING DATE: 1997-06-06  
; EARLIER APPLICATION NUMBER: 60/048,892

EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,915  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/049,019  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,970  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,972  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,916  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/049,373  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,875  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/049,374  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,917  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,949  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,974  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,883  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,962  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,963  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,877  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/048,878  
 EARLIER FILING DATE: 1997-06-06  
 EARLIER APPLICATION NUMBER: 60/070,923  
 EARLIER FILING DATE: 1997-12-18  
 EARLIER APPLICATION NUMBER: 60/092,921  
 EARLIER FILING DATE: 1998-07-15  
 EARLIER APPLICATION NUMBER: 60/094,657  
 EARLIER FILING DATE: 1998-07-30  
 NUMBER OF SEQ ID NOS: 1227  
 SOFTWARE: Patent in Ver. 2.0  
 SEQ ID NO 160  
 LENGTH: 1842  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 FEATURES:  
 NAME/KEY: SITE  
 LOCATION: (19)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (62)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (1793)  
 OTHER INFORMATION: n equals a,t,g, or c  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (1834)  
 OTHER INFORMATION: n equals a,t,g, or c  
 US-09-205-258-160  
 Alignment Scores:  
 Pred. No.: 118  
 Score: 8.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 1.72%  
 Length: 1842  
 Matches: 8  
 Conservatives: 0  
 Mismatches: 0  
 Indels: 0

DB: 4 Gaps: 0  
 US-09-744-313A-1 (1-465) x US-09-205-258-160 (1-1842)  
 QY 98 SerSerGluArgLysGluLysLys 105  
 Db 161 AGCAGTGAAGAAAGAAAGAAAGAA 184  
 RESULT 8  
 US-09-008-271A-14  
 ; Sequence 14, Application US/09008271A  
 ; Patent No. 6203979  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bandman, Olga  
 ; Hillman, Jennifer L.  
 ; Yue, Henry  
 ; Guegler, Karl J.  
 ; Corley, Neil C.  
 ; Tang, Tom Y.  
 ; Shah, Purvi  
 ; TITLE OF INVENTION: HUMAN PROTEASE MOLECULES  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
 ; STREET: 3174 Porter Dr.  
 ; CITY: Palo Alto  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94304  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette  
 ; COMPUTER: IBM Compatible  
 ; OPERATING SYSTEM: DOS  
 ; SOFTWARE: Fast-SEQ for Windows Version 2.0  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/008,271A  
 ; FILING DATE: 16-Jan-1998  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: <Unknown>  
 ; FILING DATE: <Unknown>  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Mohan-Peterson, Sheela  
 ; REGISTRATION NUMBER: 41,201  
 ; REFERENCE/DOCKET NUMBER: PF-0458 US  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 650-855-0555  
 ; TELEFAX: 650-845-4166  
 ; INFORMATION FOR SEQ ID NO: 14:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 3043 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; IMMEDIATE SOURCE:  
 ; LIBRARY: TLYMOT02  
 ; CLONE: 447484  
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 14 :  
 US-09-008-271A-14  
 Alignment Scores:  
 Pred. No.: 190  
 Score: 8.00  
 Percent Similarity: 100.00%  
 Best Local Similarity: 100.00%  
 Query Match: 1.72%  
 DB: 3  
 Length: 3043  
 Matches: 8  
 Conservatives: 0  
 Mismatches: 0  
 Indels: 0  
 Gaps: 0  
 US-09-744-313A-1 (1-465) x US-09-008-271A-14 (1-3043)  
 QY 98 SerSerGluArgLysGluLysLys 105  
 Db 291 AGCAGTGAAGAAAGAAAGAAAGAA 314

```
RESULT 9
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

Alignment Scores:
Pred. No.: 7,48e+03 Length: 152331
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.72% Indels: 0
DB: 3 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-128-155-16 (1-152331)

QY 100 GluArgLysGluLysLysGluArg 107
Db 128560 GAAAGAAAGAAAGAAAGAAAGAA 128537

RESULT 10
US-09-128-155-17/c
; Sequence 17, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; CURRENT FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650
; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 176373
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(176373)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-17

Alignment Scores:
Pred. No.: 8,58e+03 Length: 176373
Score: 8.00 Matches: 8
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0

Query Match: 1.72% Indels: 0
DB: 3 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-128-155-17 (1-176373)

QY 100 GluArgLysGluLysLysGluArg 107
Db 144583 GAAAGAAAGAAAGAAAGAAAGAA 144560

RESULT 11
US-08-110-786A-4
; Sequence 4, Application US/08110786A
; Patent No. 5443966
; GENERAL INFORMATION:
; APPLICANT: FAIRWEATHER, Neil Fraser
; APPLICANT: MAKOFF, Andrew Joseph
; TITLE OF INVENTION: Expression of tetanus toxin fragment C
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye P.C.
; STREET: 1100 No. 5443966th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/110,786A
; FILING DATE: 23-AUG-1993 1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/777,337
; FILING DATE: 29-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB90/00943
; FILING DATE: 20-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8914122.0
; FILING DATE: 20 June 1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary J. Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-134
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 161 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 30..161
US-08-110-786A-4

Alignment Scores:
Pred. No.: 128 Length: 161
Score: 7.00 Matches: 7
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 1.51% Indels: 0
DB: 1 Gaps: 0
```

US-09-744-313A-1 (1-465) x US-08-110-786A-4 (1-161)

Qy 264 ThrilleusSerProThrSer 270  
Db 118 ACGATATTATCTCCGACATCT 138

## RESULT 12

US-08-110-786A-6/c  
Sequence 6, Application US/08110786A  
Patent No. 5443966  
GENERAL INFORMATION:  
APPLICANT: FAIRWEATHER, Neil Fraser  
APPLICANT: MAKOFF, Andrew Joseph  
TITLE OF INVENTION: Expression of tetanus toxin fragment C  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon & Vanderhye P.C.  
STREET: 1100 No. 5443966th Glebe Road  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22201-4714

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/110,786A  
FILING DATE: 23-AUG-1993 1991  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/777,337  
FILING DATE: 29-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB90/00943  
FILING DATE: 20-JUN-1990

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 8914122.0  
FILING DATE: 20 June 1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Mary J. Wilson  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 117-134

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

TELEFAX: (703) 816-4100

TELEX: 200797 NLXN UR

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 161 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA

ORIGINAL SOURCE:

ORGANISM: Clostridium tetani

US-08-110-786A-6

Alignment Scores:

Pred. No.: 128

Score: 7.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 1.51%

DB: 1

Length: 161

Matches: 7

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

US-09-744-313A-1 (1-465) x US-08-110-786A-6 (1-161)

Qy 264 ThrilleusSerProThrSer 270

Db 48 ACGATATTATCTCCGACATCT 28

## RESULT 13

US-08-435-684A-43/c  
Sequence 43, Application US/08435684A  
Patent No. 5707802  
GENERAL INFORMATION:  
APPLICANT: Sandhu, Gurpreet S.  
APPLICANT: Kline, Bruce C.  
TITLE OF INVENTION: Nucleic Acid Probes for the Detection  
TITLE OF INVENTION: and Identification of Fungi  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ciba Corning Diagnostics Corp.  
STREET: 63 No. 5707802th Street  
CITY: Medfield  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02052

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 3.5 inch, 1.44 Mb storage

COMPUTER: IBM PS/2

OPERATING SYSTEM: MS-DOS 6.2

SOFTWARE: Word 6.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/435,684A

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:

NAME: Morgenstern, Arthur S.

REGISTRATION NUMBER: 28,244

REFERENCE/DOCKET NUMBER: CCD-180

TELECOMMUNICATION INFORMATION:

TELEPHONE: 508 359-3836

TELEFAX: 508 359-3885

INFORMATION FOR SEQ ID NO: 43:

SEQUENCE CHARACTERISTICS:

LENGTH: 223

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Candida glabrata specific region of 28S gene.

HYPOTHETICAL: No

ANTI-SENSE: No

US-08-435-684A-43

Alignment Scores:

Pred. No.: 174

Score: 7.00

Percent Similarity: 100.00%

Best Local Similarity: 100.00%

Query Match: 1.51%

DB: 1

Length: 223

Matches: 7

Conservative: 0

Mismatches: 0

Indels: 0

Gaps: 0

## RESULT 14

US-08-934-877A-43/c  
Sequence 43, Application US/08934877A  
Patent No. 5958693

GENERAL INFORMATION:

APPLICANT: SANDHU, Gurpreet S.

APPLICANT: KLINE, Bruce C.

TITLE OF INVENTION: DNA ISOLATION METHOD

NUMBER OF SEQUENCES: 78

CORRESPONDENCE ADDRESS:

ADDRESSEE: FOLEY & LARDNER

STREET: 3000 K Street, N.W.

CITY: Washington

STATE: D.C.

US-09-744-313A-1 (1-465) x US-08-435-684A-43 (1-223)

Qy 250 SerCysGlusSerProLysPro 256

Db 57 AGCTGGAGACTCCAGCCC 37

COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/934,877A  
FILING DATE: 22-SEP-1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/435,684  
FILING DATE: 05-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Saxe, Bern D.  
REGISTRATION NUMBER: 28,665  
REFERENCE/DOCKET NUMBER: 080394/0108  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORGANISM: Candida glabrata  
US-08-934-877A-43

Alignment Scores: 174 Length: 223  
Pred. No.: 7.00 Matches: 7  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 1.51% Gaps: 0  
DB: 2

US-09-744-313A-1 (1-465) x US-08-934-877A-43 (1-223)  
QY 250 SerCyeGluserProlysePro 256  
Db 57 AGCTGCGAGAGTCCCAAGCCC 37

RESULT 15  
US-08-871-678C-43/C  
Sequence 43, Application US/08871678C  
Patent No. 6180339  
GENERAL INFORMATION:  
APPLICANT: Sandhu, Gurpreet S, Kline, Bruce C  
TITLE OF INVENTION: Nucleic Acid Probes for the Detection and Identification  
NUMBER OF SEQUENCES: 80  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Chiron Diagnostics Corporation  
STREET: 63 No. 618039th Street  
CITY: Medfield  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02052  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS 6.2  
SOFTWARE: Word 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/871,678C  
FILING DATE: 06-JUNE-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/373,127  
FILING DATE: 13-JANUARY-1995  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/435,684  
FILING DATE: 05-MAY-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Morgenstern, Arthur S.  
REGISTRATION NUMBER: 28,244  
REFERENCE/DOCKET NUMBER: CCD-180CIP11  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 508 359-3836  
TELEFAX: 508 359-3885  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 223  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Candida glabrata specific region of 28S gene.  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-871-678C-43

Alignment Scores: 174 Length: 223  
Pred. No.: 7.00 Matches: 7  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 1.51% Gaps: 0  
DB: 3

US-09-744-313A-1 (1-465) x US-08-871-678C-43 (1-223)  
QY 250 SerCyeGluserProlysePro 256  
Db 57 AGCTGCGAGAGTCCCAAGCCC 37

Search completed: January 31, 2004, 09:06:06  
Job time: 177 secs

and is derived by analysis of the total score distribution.

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 05:12:47 ; Search time 379 Seconds  
(without alignments)  
3311.975 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 465

Sequence: 1 MYLIHFLIFRNTQGRGSF.....ELFPELNKVKQEVTSVTSWM 465

Scoring table:

OLIGO  
Xgapop 60.0, Xgapext 60.0  
Ygapop 60.0, Ygapext 60.0  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 1

Total number of hits satisfying chosen parameters: 5103490

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL-frame+ p2n.model -DRV-xlh  
-Q=/cgn2\_1/USPTO.spool/US09744313/runat\_30012004\_113611\_4581/app\_query.fasta\_1.647  
-DB=N Geneseq 19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=5 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313@cgn 1 1 0 @runat\_30012004\_113611\_4581 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Databases :

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*
- 25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	465	100.0	1992	22	AAC86397
2	456	98.1	1716	22	AA159029
3	456	98.1	2176	22	AA160815
4	456	98.1	2176	22	AA160816
5	330	71.0	3145	22	AB079518
6	307	66.0	1551	22	AA159030
7	186	40.0	779	20	AA171737
8	119	25.6	358	21	AA171737
9	80	17.2	305	25	ABX50426
10	55	11.8	402	25	ABX50611
11	38	8.2	725	24	AB577486
12	25	5.4	311	22	AAH69552
13	25	5.4	313	22	AAH72445
14	24	5.2	299	22	AAH72445
15	24	5.2	299	22	AAH73007
16	19	4.1	60	24	ABN39970
17	9	1.9	507	21	AA146735
18	9	1.9	525	21	AA146735
19	9	1.9	580	22	ABA21247
20	9	1.9	582	22	ABA21247
21	9	1.9	8774	22	ABA21246
22	9	1.9	22400	25	ABT16305
23	8	1.7	85	16	AAQ95318
24	8	1.7	234	25	ABX25823
25	8	1.7	297	23	ABV01164
26	8	1.7	317	22	AA540243
27	8	1.7	317	22	AA540243
28	8	1.7	317	22	AA102161
29	8	1.7	345	22	AA134724
30	8	1.7	345	25	ABX57712
31	8	1.7	397	23	ABV31504
32	8	1.7	397	23	ABV40473
33	8	1.7	400	21	AAF15721
34	8	1.7	400	22	AA540122
35	8	1.7	400	22	AA540122
36	8	1.7	420	22	AA535219
37	8	1.7	421	21	AA174769
38	8	1.7	421	24	ABN26460
39	8	1.7	423	23	ABV03089
40	8	1.7	431	23	ABV20749
41	8	1.7	431	23	ABV22006
42	8	1.7	431	23	ABV25932
43	8	1.7	431	23	ABV27839
44	8	1.7	440	23	ABV12258
45	8	1.7	450	22	AA125723

ALIGNMENTS

RESULT 1  
AAC86397  
ID AAC86397 standard; DNA; 1992 BP.

XX AAC86397;

DT 21-MAR-2001 (first entry)

XX SNEKN DNA #1.

KW SNEKN; human; sorting nexin; inflammation; asthma; allergy; AIDS;  
KW neurological disorder; gastrointestinal; smooth muscle cell;  
KW cancer; gene therapy; ds.

OS Homo sapiens.

XX WO200073334-A2.



XX 07-DEC-2000.  
 XX 26-MAY-2000; 2000WO-US14831.  
 XX 27-MAY-1999; 99US-0136740.  
 XX 16-JUN-1999; 99US-0139566.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Yue H, Tang YT, Azimzai Y;  
 XX WPI; 2001-041141/05.  
 XX Novel Sorting Nexin polypeptides and polynucleotides useful for  
 XX diagnosing and treating disorders associated with their expression e.g.  
 XX PT autoimmune disorders, smooth muscle cell disorders and cell  
 XX PT proliferative disorders  
 XX Claim 5; Page 82-83; 84pp; English.  
 XX The present invention relates to human sorting nexin (SNEXN).  
 CC Compositions containing SNEXN or agonists of SNEXN are useful for  
 CC treating a disease or condition associated with decreased expression of  
 CC functional SNEXN and compositions containing antagonists of SNEXN are  
 CC useful for treating a disease or condition associated with  
 CC overexpression of functional SNEXN. These can be immune  
 CC disorders for example inflammation, asthma, allergy, and AIDS,  
 CC neurological disorders, gastrointestinal disorders, smooth muscle cell  
 CC disorders, cancers and others. The SNEXN may also be used for somatic  
 CC or germline gene therapy.  
 XX Sequence 1992 BP; 671 A; 308 C; 372 G; 641 T; 0 other;  
 SQ

Alignment Scores:  
 Pred. No.: 0 Length: 1992  
 Score: 465.00 Matches: 465  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0

US-09-744-313A-1 (1-465) x AAC86397 (1-1992)

Qy 1 MetTyrLeuLeuHisPheCysLeuLeuPheArgAsnThrGlnLysArgGlyGluSerPhe 20  
 Db 116 ATGACTTGATACATTTTGTAAATATTCAGGAACACACAGAAAAGGGGAGAAATCATTT 175  
 Qy 21 GlyTleSerArgIleGlySerLysLysLysGlyValPheLysSerThrThrMetGluGly 40  
 Db 176 GGAATCAGCAGATAGGTAGCAGAAATTAAGGAGATATTCAGAAAGTACCACAAATGGAGGGA 235  
 Qy 41 AlaMetLeuProAsnTyrGlyValAlaGluGlyGluAspPheIleGluGluGlyIle 60  
 Db 236 GCTATGTTGCTTAATATGTTAGTGTAGCTGAAGGTGAAGTATTTATGAGAGAGTATT 295  
 Qy 61 ValValMetGluAspAspSerProValGluAlaValSerThrProAsnThrProArgAsn 80  
 Db 296 GTTGTATGAGAGATGATTCCTCAGTGGAGGCTGTGAGCACCTAATACCTCCCGAAC 355  
 Qy 81 LeuAlaLalTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 100  
 Db 356 CTTGCTGATGAGAAATAGCATTCATATGATAGTATTTTGGAGTATCCTCTCTGAA 415  
 Qy 101 ArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArgAsnAspArg 120  
 Db 416 AGAAGAGAGAGAAAGAAAGAAATTCCTGTGTGTTGTATGATGTGAAGAAATATAGTA 475  
 Qy 121 ArgAlaValGlyHisGluProGluHisTrpSerValTyrArgArgTyrLeuGluPheTyr 140  
 Db 476 AGAGCAGTGTGACACGACGCTGACATTTGCTCTCTATAGAGATATCTTGAATCTAT 535  
 Qy 141 ValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSer 160

Db 536 GTACTTGAATCAAACTAAAGAAATTCATGTCATTTCCGTATGCCAGCTTCCTCTT 595  
 Qy 161 LysArgIleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGln 180  
 Db 596 AAGAGGATCATTCGCCCAAAATTAATGAATCTTAAGTCAAGAGGGAAGAGTCCAA 655  
 Qy 181 GluTyrLeuGlnLysLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaApp 200  
 Db 656 GAATATCTACAGAAACTTCTGCAGCATCCAGAACTGAGTAATAGTCAACTCTCTGGCAGAC 715  
 Qy 201 PheLeuSerProAsnGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsn 220  
 Db 716 TTTCTTTCCCTAATGTCGGGAAACACAAATTTCTTGATAAGATATACAGATGTAAT 775  
 Qy 221 LeuGlyLysLysLysSerValProGlyLysLeuMetLysGluLysGlnHisLeu 240  
 Db 776 CTTGGGAAATTAATAATCTGTTCTTGGAAATTAATGAAAGAAAGGTGAGCTTTG 835  
 Qy 241 GluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArg 260  
 Db 836 GAACCTTTTATCATGAATTTTCAATTAATCTTGTGAGTCTCCAAAGCTTAACCAAGTAGA 895  
 Qy 261 ProGluLeuThrIleLeuSerProThrSerGluAsnLysLysLysLeuPheAsnAspLeu 280  
 Db 896 CCAGAACTGACCATTTCTCAGCCCTACTTCAGAAAACACAAAGAGCTTTTCAATGATCTG 955  
 Qy 281 PheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPhe 300  
 Db 956 TTTAAAAATTAATCAAAACCGTGTGAAATACAGAGAGAAAGCAAAATCAAGATTTATTT 1015  
 Qy 301 MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal 320  
 Db 1016 ATGAGGTGATGACTGTAGAGGAGTCTATGATTAATCTGATGATGATGAGGCGGTAGTT 1075  
 Qy 321 PheGlnValProAspTrpLeuHisHisLeuLeuMetGlyThrArgIleLeuPheLysAsn 340  
 Db 1076 TTCCAGGTTCCTGACTGGCTTCATCATCTCTTAATGGAACTCGAATCTCTTTAAAAAC 1135  
 Qy 341 ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGlnLeuPheGlnGlu 360  
 Db 1136 ACCCTGGAAATGATATCTACTTCTTCAAGTGTAACTAGAACAGCTATTTTCAGAG 1195  
 Qy 361 HisArgLeuValSerLeuLeuThrLeuLeuArgAspAlaIlePheCysGluAsnThrGlu 380  
 Db 1196 CACGTTTGGTCTCCTCATCATCACTCTCTCAGATGCTATATCTGTGAAACACTGAA 1255  
 Qy 381 ProArgSerLeuGlnAspLysGlnLysGlyValLysGlnThrPheGluGluMetMetAsn 400  
 Db 1256 CTTGCTCTCTCCAGATAGCAAAAGAGCAAAACAGACTTTTGAAGAAATCATGAAT 1315  
 Qy 401 TyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArg 420  
 Db 1316 TACATTCAGATCTGTTAGTCAAGTATTGTCAGAGAAACCAAGTATGAAGCATCAGA 1375  
 Qy 421 LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu 440  
 Db 1376 CTTCTGTTGATGCTTACAGCAACCACTACTCAACAGCAGCTGACTATGTTTATTG 1435  
 Qy 441 AspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSer 460  
 Db 1436 GACATTGTGATACAGGAATCTGTTTCCAGAGCTCAATAGGTACAAAGGAAGTACCTCT 1495  
 Qy 461 ValThrSerTrpMet 465  
 Db 1496 GTGACATCTGGATG 1510

RESULT 2  
 AAI59029  
 ID AAI59029 standard; cDNA; 1716 BP.  
 XX  
 AC AAI59029;  
 XX

DT	22-OCT-2001	(first entry)	Db	173	TTTCGGACACACAGAAAAGGGGAGAAATCATTTGGATCAGCAGAAATAGGTAGCAAAATT	232
XX	Human polynucleotide SEQ ID NO 1232.		Qy	30	LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla	49
XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		Db	233	AAAGAGTATTCAAAGTACCACAAATGGAGGAGCTATGTTGCTTAATATGTTAGTCT	292
XX	Peripheral nervous system; neuropathy; central nervous system; CNS;		Qy	50	GluGlyGluAspPheIleGluGlyIleValMetGluAspSerProVal	69
XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		Db	293	GAAGTGAAAGATGATTTATTGAAGAAGGTATGTTGAATGAAGATGATTCTCCAGTG	352
XX	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		Qy	70	GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTyrLysSerIlePro	89
XX	chemokine; thrombolytic; drug screening; arthritis; inflammation;		Db	353	GAGGCTGTGAGCACACCTTAATATCTCCCGAAACCTTGTGCAATGGAANATAGCATTC	412
XX	leukaemia; ss.		Qy	90	TyrValAspPhePheGluAspProSerSerGluArgLysGluLysGluArgIlePro	109
OS	Homo sapiens.		Db	413	TATGTAGACTTTTGTGAGGATCCCTCTCTGAAGAGGAGAGAAAAGAAATTCCT	472
XX	WO200153312-A1.		Qy	110	ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis	129
XX	26-JUL-2001.		Db	473	GTGTTTGTATTGATGTTGAAAGAAATGATAGAGAGAGTTCGACACGAGCTGAACAT	532
XX	26-DEC-2000; 2000WO-US34263.		Qy	130	TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe	149
XX	21-JAN-2000; 2000US-0488725.		Db	533	TGCTCTGTCTATAGAAAGATATCTTGAATTCATGTTGAATCAAAACCTAACAGATT	592
XX	25-APR-2000; 2000US-0552317.		Qy	150	HisGlyValPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr	169
XX	09-JUL-2000; 2000US-0598042.		Db	593	CATGGTGCAATTCCTGATGCCAGCTTCCTTCTTAGAGGATCATTTGGCCCCAAAATAT	652
XX	19-JUL-2000; 2000US-0620312.		Qy	170	GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHis	189
XX	03-AUG-2000; 2000US-0653450.		Db	653	GAATTCCTAAAGTCAAAGAGGAGGAGTTCCAAGAAATATCTACAGAAATCTTCGAGCAT	712
XX	14-SEP-2000; 2000US-0662191.		Qy	190	ProGluLeuSerAsnSerGlnLeuAlaAspPheLeuSerProAsnGlyGlyGluThr	209
XX	19-OCT-2000; 2000US-0693036.		Db	713	CCAGAACTCAGTAATAGTCAACTTCCTGGCAGACTTCTTCTCCCTTAATGCTGGGAAACA	772
XX	29-NOV-2000; 2000US-0727344.		Qy	210	GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysValSerValPro	229
XX	(HYSE-) HYSEQ INC.		Db	773	CAATTTCTTGATAGATACCTACAGATGTAATCTTGGGAAATATAAAATCTGTCTCT	832
XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		Qy	230	GlyLysLeuMetLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn	249
XX	PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		Db	833	GGAATACTAATGAAGAGAAAGGTGAGCTTTGGAACCTTTTATCATGATTTTCAATAT	892
XX	PI Zhao QA, Zhou P, Goodrich R, Dmanac RT;		Qy	250	SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr	269
XX	WPI; 2001-442253/47.		Db	893	TCITGTGAGTCTCAAAGCTTAAACCAAGTACAGACAGAACTGACCATTTCTCAGCCCTACT	952
XX	P-PSDB; AAM39873.		Qy	270	SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu	289
XX	Novel nucleic acids and polypeptides, useful for treating disorders		Db	953	TCAGAAAACAAACAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTGTGAA	1012
XX	such as central nervous system injuries -		Qy	290	AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal	309
XX	Claim 1; SEQ ID NO 1232; 10078pp; English.		Db	1013	AATACAGAGAGAAAGAAAATCAGAAATATTTTATGAGGTGATGACTGTAGAGAGATC	1072
XX	The invention relates to human nucleic acids (AAI57798-AAI61369) and		Qy	310	TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspThrLeuHisHis	329
XX	the encoded polypeptides (AAM39642-AAM42213) with nootropic,		Db	1073	TATGATTAATGAGTATGATGATGAGCGGTAGTTTCCAGGTCTCTGCTGCTTCATCAT	1132
XX	immunosuppressant and cytostatic activity. The polynucleotides are useful		Qy	330	LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr	349
XX	in gene therapy. A composition containing a polypeptide or polynucleotide		Db	1133	CTCTTAATGGAACTCGAATCTCTCTTTAAATAACCCCTGGAATGATATGATGATGAT	1192
XX	of the invention may be used to treat diseases of the peripheral nervous		Qy	350	LeuGlnCysLysLysLeuGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu	369
XX	system, such as peripheral nervous injuries, peripheral neuropathy and		Db	1193	CTTCAGGTAAACTAGAACAGCTATTTCAGGAGCACCCTTGGTCTTCATCATACACTT	1252
XX	localised neuropathies and central nervous system diseases, such as		Qy	370	LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys	389
XX	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		Db	1253	CTCAGAGATCTATATCTCTGTGAAACACTGAACCTCGCTCTCTCCAGAGTAAGCAAAA	1312
XX	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		Qy			
XX	utilisation of the activities such as: Immune system suppression,		Db			
XX	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		Qy			
XX	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		Db			
XX	assays for receptor activity, arthritis and inflammation, leukaemias and		Qy			
XX	C.N.S disorders.		Db			
XX	Note: The sequence data for this patent did not form part of the printed		Qy			
XX	specification.		Db			
XX	Sequence 1716 BP; 572 A; 297 C; 347 G; 500 T; 0 other;		Qy			
XX	Alignment Scores:		Db			
XX	Pred. No.: 0	Length: 1716	Qy			
XX	Score: 456.00	Matches: 456	Db			
XX	Percent Similarity: 100.00%	Conservative: 0	Qy			
XX	Best Local Similarity: 100.00%	Mismatches: 0	Db			
XX	Query Match: 98.06%	Indels: 0	Qy			
XX	DB: 22	Gaps: 0	Db			
XX	US-09-744-313A-1 (1-465) x AAI59029 (1-1716)		Qy			
XX	10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29		Db			

QY 390 GlyValAspGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuVallyscys 409  
 Db 1313 GAGCAAAACAGACTTTTGAAGAAATGATGATTAATCCAGATCTGTAGTCAAGTGT 1372  
 QY 410 IleGlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnPro 429  
 Db 1373 ATTGGTGAAGAAACCAAGATGAAAGCATCAGACTTCTCTTTGATGGCTTACAGCAACA 1432  
 QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnLeuPhePro 449  
 Db 1433 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTTGATACAGGAATCTGTTCOA 1492  
 QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
 Db 1493 GAGCTCAATAGGTACAAAGAGTACTCTGTGACATCTTGGATG 1540

## RESULT 3

AAI60815/c  
 ID AAI60815 standard; cDNA; 2176 BP.

XX AAI60815;

DT 22-OCT-2001 (first entry)

XX Human polynucleotide SEQ ID NO 4804.

XX Human; neurotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.

XX Homo sapiens.

OS WO200153312-A1.

PN 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPT; 2001-442253/47.

XX P-PSDB; AAM41659.

XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -

XX Claim 1; SEQ ID NO 4804; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemia and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX SQ Sequence 2176 BP; 684 A; 427 C; 353 G; 712 T; 0 other;

## Alignment Scores:

Pred. NO.: 0 Length: 2176  
 Score: 456.00 Matches: 456  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 98.06% Indels: 0  
 DB: 22 Gaps: 0

US-09-744-313A-1 (1-465) x AAI60815 (1-2176)

QY 10 PheArgSerThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29  
 Db 2022 TTTCCGAACACACAGAAAGGGGAGAAATCAATTTGGAAATCAGCAGAAATAGGTAGCAAAAT 1963  
 QY 30 LysGlyValPheLysSerThrThrMetGluClyValMetLeuProAsnTyrGlyValAla 49  
 Db 1962 AAAGAGATTTCAAAAGTACCACAAATGGAGGAGCTATGTTCCTAATATATGCTGTAGCT 1903  
 QY 50 GluGlyGluPheAspPheIleGluGluGlyValValMetGluAspSerProVal 69  
 Db 1902 GAAGTCTCAGATGATTTTATTGAAGAGGTATTGTTGTAATGAAGATGATCTCCAGTG 1843  
 QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaLalaTrpLysIleSerIlePro 89  
 Db 1842 GAGGCTGTGACACACCTAATCTCCCGAAACCTTCTCATGGAATAATAGCATTTCCA 1783  
 QY 90 TyrValAspPhePheGluAspProSerSerGluArgLysGluLysValGluArgIlePro 109  
 Db 1782 TATGTAGACTTTTGTAGGATCCCTCTCTGAAGAGGAGGAGAAAGAAAGAAATTCCT 1723  
 QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
 Db 1722 GTGTTTGTATTGTTGAAAGAAATGATAGAGACAGCTTGACACAGAGCTGAACAT 1663  
 QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
 Db 1662 TGGTCTGTCTATAGAGATATCTTGAATTCCTATGATCTTGAATCAAAACTAACAGAAATTT 1603  
 QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerIleValIleGlyProLysAsnTyr 169  
 Db 1602 CATGGTGCATTTCTGTATGCCAGCTTCTTCTAAGAGGATCATTTGCCCCCAAAATAT 1543  
 QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis 189  
 Db 1542 GAATCTTAAAGTCAAGAGGAGAGAGTTCAGAGATATCTACAGAACTCTTGACAT 1483  
 QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
 Db 1482 CCAGAGCTGACTAATAGTCAACTTCTGCGACAGCTTTCTTCCCTAATAGTGGGGAACA 1423  
 QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValPro 229  
 Db 1422 CAATTTCTTCAATAAGTACTACACAGATGTAATCTTGGGAAATATTAATAATCTGTTCCT 1363  
 QY 230 GlyLysLeuMetLysGluLysGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
 Db 1362 GGAAACTTAATGAAGAGAGAGAGTTCAGACTTTTGAACCTTTTATCATGAATTCATTAAT 1303  
 QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
 Db 1302 TCTTGTGAGTCTCCAAAGCCTTAACCAAGTAGACCAAGTACTGACCAATTCCTCAGCCCTACT 1243

QY 270 SerGluAsnAspLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
 Db 1242 TCAGAAACACACAGAGCTTTTCATGATCTGTTTAAATAATAGCAACCGTCTGAA 1183  
 QY 290 AsnThrGluArgGlnAsnGlnAsnThrPheMetGluValMetThrValGluGlyVal 309  
 Db 1182 AATACAGAGAGAGAGCAAAATCAGAAATATTTTATGAGAGGTGATGACTGTAGAGAGCTC 1123  
 QY 310 TyrAspTyrLeuMetTyrValGluArgValValPheGlnValProAspTyrLeuHis 329  
 Db 1122 TATGATTACCTGATGATGATGAGCGGTGAGTTTCCAGGTTCCTGAGCTGGCTTCATCAT 1063  
 QY 330 LeuLeuMetGlyThrArgGlnLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
 Db 1062 CTCCTTAATGGAACTCGAATCTCTTTAAACACCCCTGGAAATGATGATGATGAT 1003  
 QY 350 LeuGlnCysGlyLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeu 369  
 Db 1002 CTTCAAGTGAATCAAGACAGCTATTTCCAGGAGCACCGTTTGGTCTCACTCATACACTT 943  
 QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
 Db 942 CTCAGAGATGCTATATTTCTGAAACACACTGAACTGCTCTCTCAAGATAAGCAAAA 883  
 QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
 Db 882 GAGCAAAACAGACTTTTGAAGAAATGATGAAATATTCAGATCTGTTAGTCAAGTGT 823  
 QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
 Db 822 ATTGGTGAAGAACCAAGATGAAAGCATCAGACTCTCTGTTGATGGCTTACAGCAACA 763  
 QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
 Db 762 GTACTCAACAAGCAGCTGACTATGTTTATTTGGAATGTTGATGATGATGATGATGAT 703  
 QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
 Db 702 GAGCTCAATAGGTACAAAGGAAGTACCTCTGTGACATCTGGATG 655

RESULT 4  
 AA160816/c  
 ID AA160816 standard; cDNA; 2176 BP.  
 XX AC AA160816;  
 XX XX  
 XX 22-OCT-2001 (first entry)  
 XX DE Human polynucleotide SEQ ID NO 4805.  
 XX KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW KW leukaemia; ss.  
 XX OS Homo sapiens.  
 XX XX  
 XX PN W0200153312-A1.  
 XX XX  
 XX PD 26-JUL-2001.  
 XX XX  
 XX PF 26-DEC-2000; 2000WO-US34263.  
 XX XX  
 XX FR 21-JAN-2000; 2000US-0488725.  
 XX FR 25-APR-2000; 2000US-0552317.  
 XX PR 09-JUL-2000; 2000US-0598042.  
 XX PR 19-JUL-2000; 2000US-0620312.  
 XX PR 03-AUG-2000; 2000US-0653450.  
 XX PR 14-SEP-2000; 2000US-0662191.  
 XX PR 19-OCT-2000; 2000US-0693036.  
 XX PR 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX WPI; 2001-442253/47.  
 DR P-PSDB; AAM41660.  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 PT Claim 1; SEQ ID NO 4805; 10078pp; English.  
 XX The invention relates to human nucleic acids (AA157798-AA161369) and  
 CC the encoded polypeptides (AAW38642-AAW42213) with nontropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemia and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX SQ Sequence 2176 BP; 684 A; 427 C; 353 G; 712 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0 Length: 2176  
 Score: 456.00 Matches: 456  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 98.06% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-744-313A-1 (1-465) x AA160816 (1-2176)  
 QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyLysSerArgIleGlySerLysIle 29  
 Db 2022 TTTCCGAAACACACAGAAAGGGGAGATCATTTTGAATCAGCAGATAGTAGTACAAAT 1963  
 QY 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
 Db 1962 AAAGGAGTATTTCAAAAGTACCACATGGAGGGAGCTATGTTGCTTAATTATGTGTAGCT 1903  
 QY 50 GluGlyGluAspAspPheIleGluGluGlyValValMetGluAspAspSerProVal 69  
 Db 1902 GAAGTGAAGATGATTTTATTTATGAGAGGATTTGTTGTAATGGAAGATGATTTCCAGTG 1843  
 QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro 89  
 Db 1842 GAGCTGTGAGCACACCTAATCTCCCGAAACCTTGCTGCTGAAATATGATTTCCA 1783  
 QY 90 TyrValAspPheGluAspProSerSerGluArgLysGluLysGluArgIlePro 109  
 Db 1782 TATGTAGACTTTTGTGAGTATCCCTCTCTGAAAGGAGGAGAGAAAGAAAGAAATTCCT 1723  
 QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
 Db 1722 GTGTTTTGTATGATGTTGAAGAAATGATAGAGAGCAGTGTGACACAGCCTGAACAT 1663  
 QY 130 TrpSerValTyrArgAspTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
 Db 1662 TGGTCTGTCTATAGAGATATCTTGATTTCTATGATCTTGAATCAAACTAACAGATTT 1603  
 QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169

DB 1602 CATGGTCATTCCTGATGCCAGCTCTCTCTAGAGGATCATTTGGCCCCAAAATAT 1543  
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis 189  
DB 1542 GAATTCCTTAAGTCAAGAGGGAAGAGTTCAGGAATATCTACAGAAACTTCTGCAGCAT 1483  
QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
DB 1482 CCAGAACTCAGTAATAGTCAACTCTGGCAGACTTCTCTCCCTTAATGGTGGGAAACA 1423  
QY 210 GlnPheLeuAspLysLysLeuProAspValAsnLeuGlyLysLysLysLysSerValPro 229  
DB 1422 CAATTCCTTGATTAAGATATCTACAGATGTAATCTTGGGAAATATTAATCTGTCTCT 1363  
QY 230 GlyLysLeuMetLysGlyGlnHisLeuGluProPheLeuMetAsnPheLeuLeu 249  
DB 1362 GGAAACTAATGAAGAGAAAGTCCAGCATTTGGAACTTTTATCATGAATTCATTAAT 1303  
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrLysLeuSerProThr 269  
DB 1302 TCTTGTGAGTCTCCAAAGGCTAAACCAAGTAGACCAAGAACTGACCATTCTCAGCCCTACT 1243  
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
DB 1242 TCAGAAACCAACAGAAAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAA 1183  
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
DB 1182 AATACAGAGAAAGCAAAATCAGATTTATTTATGGAGTGATGATCTAGAGAGTCT 1123  
QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrPheHisHis 329  
DB 1122 TATGATTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1063  
QY 330 LeuLeuMetGlyThrArgLysLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
DB 1062 CTCTTAATGGAACTCGAATCTCTTTAAACACACCTCGAATGATGATGATGATGAT 1003  
QY 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeu 369  
DB 1002 CTTGAGTGTAACTAGACAGCTATTTTCAGGACACCGTTTGTCTCTACTATACATCT 943  
QY 370 LeuArgAspAlaLysPheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
DB 942 CTCAGAGATGCTATATCTCTGAAACACCTGAACTCGCTCTCTCCAGATAAGCAAAA 883  
QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrLysLeuProAspLeuLeuValLysCys 409  
DB 882 GGAGCAAAACAGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGAT 823  
QY 410 IleGlyGluGlnThrLysTyrGluSerLysLeuLeuPheAspGlyLeuGlnPro 429  
DB 822 ATTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGTATGCTTACAGCAACA 763  
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspLysValLysGlnLeuPhePro 449  
DB 762 GTATCTCAACAGAGCTGATCTATGTTTATTTATTTGACATTTGATGATGATGATGAT 703  
QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
DB 702 GAGCTCAATAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 655

RESULT 5

ABQ79518  
ID ABQ79518 standard; cDNA; 3145 BP.  
XX  
AC ABQ79518;  
XX

DT 25-NOV-2002 (first entry)

DE cDNA encoding a protein similar to human sorting nexin.

XX

KW Atherosclerosis; antiarteriosclerotic; marker; cardiovascular;  
KW SSH 6; gene; human; nexin; ss.  
XX

OS Homo sapiens.

XX Key Location/Qualifiers  
FH 128..2968  
CDS /\*tag= a

XX W0200262839-A2.

XX 15-AUG-2002.

XX 05-FEB-2002; 2002WO-EP01327.

XX 07-FEB-2001; 2001EP-0200439.

XX (UTMA-) UNIV MAASTRICHT.

XX Daemen MWAP, Cleutjens CBJM, Zaman GJR;

XX WPI; 2002-643400/69.

XX P-PSDB; ABB81193.

XX Use of a polynucleotide differentially expressed in ruptured and stable  
XX atherosclerotic plaques as a marker for atherosclerosis, useful in  
XX treating, diagnosing or preventing atherosclerosis -

XX Example 6; Page 35-37; 44pp; English.

XX The invention relates to the use of a polynucleotide differentially  
XX expressed in ruptured and stable atherosclerotic plaques as a marker  
XX for atherosclerosis, where the polynucleotides can be selected from the  
XX sequences shown in ABQ79517-19. The polynucleotides are useful as a  
XX marker of atherosclerosis, which may be used: (i) in the diagnosis,  
XX prevention and treatment of atherosclerosis; (ii) as serum/plasma markers  
XX to screen patients at risk for plaque instability to evaluate the effects  
XX of other treatments; (iii) in the preparation of vector molecules for the  
XX expression of the encoded protein in host cells; and (iv) in the  
XX identification of functional targets or analogues of the gene. The  
XX polynucleotides, the encoded proteins or antibodies against the proteins  
XX may be used to target other therapeutics to an unstable plaque.  
XX Modulation of the expression of the polynucleotide can increase plaque  
XX stability and therefore inhibit the progression of atherosclerotic  
XX cardiovascular disease. Modulators may be used to prepare pharmaceuticals  
XX for atherosclerotic disorders. The present sequence represents a cDNA  
XX encoding a protein similar to human sorting nexin.

XX SQ Sequence 3145 BP; 985 A; 570 C; 659 G; 931 T; 0 other;

Alignment Scores:

Pred. No.: 0 Length: 3145  
Score: 330.00 Matches: 430  
Percent Similarity: 99.54% Conservative: 0  
Best Local Similarity: 99.54% Mismatches: 1  
Query Match: 70.97% Indels: 2  
DB: 24 Gaps: 0

US-09-744-313A-1 (1-465) x ABQ79518 (1-3145)

QY 35 SerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGluGlyGluAspAsp 54  
DB 1673 AGTACCACATGAGGGAGCTATGTCCTAAATATGCTAGCTGAAGGTGAAGATGAT 1732

QY 55 PheIleGluGlyLysValValMetGluAspSerProValGluAlaValSerThr 74  
DB 1733 TTTATTGAAGAGGTATTGTTGTAATGAAGATGATCTTCAGTGGAGGCTGTGAGCACA 1792

QY 75 ProAsnThrProArgAsnLeuAlaAlaTrpLysLysSerIleProTyrValAspPhePhe 94  
DB 1793 CCTAATACCTCCCGAAACCTTGTGCTGCAATGGAATATTAGCATTCATATGATAGACTTTT 1852

QY 95 GluAspProSerSerGluArgLysGluLysLysGluArgLysProValPheCysIleAsp 114

Db	1853	GAGGATCCCTCCTCTGAAGAGGAGGAGAAAAAGAAAGAAATTCCTGTTGTTTTGTTATGAT	1912
Qy	115	ValGluArgAsnAspArgAlaValAlGlyHisGluProGluHisIstpSerValTyrArg	134
Db	1913	GTTGAPAGAATGATAGAGGACAGTGGACACGAGCCTGACATTTGGTCTGCTATAGA	1972
Qy	135	ArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAlaPhePro	154
Db	1973	AGATATCTTGAATTCATGTACTTGAATCAAACTAAACCTAACAGAAATTCATGCTGCAATTCCT	2032
Qy	155	AspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLysSer	174
Db	2033	GATGCCAGCTTCCTCTAAGAGGATCATTTGGCCCCAAAAATATGAAATTCCTTAAAGTCA	2092
Qy	175	LysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHisProGluLeuSerAsn	194
Db	2093	AAGAGGGAAGAGTCCAGGAATATCTACAGAACTTCTGCAGCATCCAGAACTGAGTAAT	2152
Qy	195	SerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLys	214
Db	2153	AGTCAACTTCTGGCAGACTTCTTCCCTTAATGGTGGGAAACACAAATTCCTTGATAAG	2212
Qy	215	IleLeuProAspValAsnLeuGlyLysIleIleLysSerValProGlyLysLeuMetLys	234
Db	2213	ATACTACAGATGTAAATCTTTGGGAAATATATAAATCTGTTCTCGTGAACCTAAATGAAA	2272
Qy	235	GluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerPro	254
Db	2273	GAGAAAGGTGAGCAATTTGGAACTTTTATCATGAATTTCAATTAATCTTGAGTCTCCA	2332
Qy	255	LysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLys	274
Db	2333	AAGCCTAAACCAAGTAGACCAGAACTGCACATCTCTAGCCCTACTTCAGAAAAACAACAAG	2392
Qy	275	LysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgalagLuanThrGluLysLys	294
Db	2393	AAGCTTTTCAATGATCTGTTTAAATAATATGCACACCGCTGCGAAATACAGAGAGAAAG	2452
Qy	295	GlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMet	314
Db	2453	CAMATCAGAAATTAATTTATGGAGTGATGACTGTAGAAGAGTCTATGATTAACCTGATG	2512
Qy	315	TyrValGlyArgValValPheGlnVal-ProAspTrpLeuHisLeuLeuMetGlyLys	334
Db	2513	TATGTAGACGGGTAGTTTTCCAGAT-TCCTGACTGGCTTCATCATCTCTTAATGGGAAC	2571
Qy	334	ArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLe	354
Db	2572	TCGAATCCTCTTTAAAAACACCCCTGGAAATGTACTGATTAATCTTCTCAGTGTAACT	2631
Qy	354	uGluGlnPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIle	374
Db	2632	AGAACCTATTTTCAGAGACACCGTTTGGTCTCACTCATAACCTCTCAGAGATGCTAT	2691
Qy	374	ePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThr	394
Db	2692	ATTCGTGAAAAACACTGAACCTCGCTCTCTCCAGATAAGCAAAAAGAGCAAAAACAGAC	2751
Qy	394	rPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluThr	414
Db	2752	TTTTTGAAGAAATGATGAATTACATTCAGATCTGTTAGTCAAGTGTATTGGTCAGGAAC	2811
Qy	414	rLysTyrGluSerIleArgIleuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGln	434
Db	2812	CAAGTATGAAGCATCAGACTTCTGTTGTGATGGCTTACAGCAACCACTACTCAACAGCA	2871
Qy	434	nLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGluLeuAsnLysVal	454
Db	2872	GCTGACCTATGTTTTATTGGACATTTGTATGATACAGGAACCTGTTTCCAGAGCTCAATAGGT	2931
Qy	454	lGlnLysGluValThrSerValThrSerTrpMet	465

Db	2932	ACAAAGGAGTACTCTGTGACATCTTGATG	29365
RESULT 6			
AAI59030			
ID	AAI59030	standard; cDNA; 1551 BP.	
XX	XX		
AC	AAI59030;		
XX			
DT	22-OCT-2001	(first entry)	
XX			
DE	Human polynucleotide SEQ ID NO 1233.		
XX			
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukemia; ss.		
OS	Homo sapiens.		
XX			
FN	WO200153312-A1.		
XX			
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000WO-US34263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	03-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
PR	29-NOV-2000; 2000US-0727344.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang V, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
XX			
DR	WPI; 2001-442253/47.		
DR	P-PSDB; AAM39874.		
XX			
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
XX			
PS	Claim 1; SEQ ID NO 1233; 10078pp; English.		
XX			
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and		
CC	the encoded polypeptides (AAM38642-AAM42213) with nootropic,		
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: immune system suppression,		
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		
CC	assays for receptor activity, arthritis and inflammation, leukaemias and		
CC	C.N.S disorders.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification.		
XX			
SQ	Sequence 1551 BP; 527 A; 258 C; 317 G; 449 T; 0 other;		
Alignment Scores:			
Pred. No.:	1.78e-307	Length:	1551
Score:	307.00	Matches:	307
Percent Similarity:	100.00%	Conservative:	0





US-09-744-313A-1 (1-465) x AAZ17379 (1-779)	
QY 263 LeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLys 282	
DB 80 CTGACCATTCCTCAGTCTCTACTTTCAGAAAACAACAAGAGCTTTCAATGATCTGTTAAA 139	
QY 283 AsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGlu 302	
DB 140 AATAATGCAACCGTCTCTGAAATATACAGAGAGAAAGCAAAATCAGAAATATTATTTATGGAG 199	
QY 303 ValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGln 322	
DB 200 GTGATGACTGTAGAAGAGTCTATGATTACCTGATGTATGTAGGACGGGTAGTTTTCAG 259	
QY 323 ValProAspTyrLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeu 342	
DB 260 GTTCTGACTGGCTTCATCTCTTAATGGGACCTCGATCTCTTTAAAAACACCTG 319	
QY 343 GluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArg 362	
DB 320 GAAATGTATACTGATTACTATCTTCAGTGTAAACTAGAACAGCTATTTCAGGAGCACCGT 379	
QY 363 LeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArg 382	
DB 380 TTGGTCTCACTATAACACTTCTCAGAGATGCTATATTCTGTGAAAACACTGAACCTGCG 439	
QY 383 SerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIle 402	
DB 440 TCTCTCAGATTAAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGATACATT 499	
QY 403 ProAspLeuLeuValLysCysIleGlyGluThrLysTyrGluSerIleArgLeuLeu 422	
DB 500 CCAGATCTGTAGTCAAGTGTATTGTGGAAGAACCAAGTATGGAAGCATCAGACTTCG 559	
QY 423 PheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIle 442	
DB 560 TTTGATGGCTTACAGCAACCGACTCTCAACAGCAGCTGACTTATGTTTATTGGACATT 619	
QY 443 ValIleGlnGluLeuPhe 448	
DB 620 GTGATCAGGAAGTGT 637	
RESULT 8	
AA02755	
ID AAC02755 standard; cDNA; 358 BP.	
XX AC AAC02755;	
DE 06-OCT-2000 (first entry)	
XX Human secreted protein 5' EST, SEQ ID NO: 2753.	
XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;	
KW gene therapy; chromosome mapping; ss.	
XX Homo sapiens.	
XX BP1033401-A2.	
XX 06-SEP-2000.	
PF 21-FEB-2000; 2000EP-0200610.	
XX 26-FEB-1999; 99US-0122487.	
XX (GEST ) GENSET.	
XX Dumas Milne Edwards J, Duclert A, Giordano J;	
XX WPI; 2000-500381/45.	
XX P-PSDB; AAG02749.	
XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for	
PT	

PT obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for	
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -	
PS Claim 1; SEQ ID 2753; 71pp + CD-ROM; English.	
XX The present sequence is one of a large number of 5' ESTs derived from	
CC mRNAs encoding secreted proteins. An ORF has been identified within the	
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs	
CC derived from 30 different tissues. EST sequences usually correspond	
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are	
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not	
CC well suited for isolating cDNA sequences derived from the 5' ends of	
CC mRNAs and even in those cases where longer cDNA sequences have been	
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from	
CC cDNAs with intact 5' ends and can therefore be used to obtain full length	
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,	
CC gene therapy and chromosome mapping procedures. They are used to obtain	
CC upstream regulatory sequences and to design expression and secretion	
CC vectors.	
XX SQ Sequence 358 BP; 125 A; 63 C; 69 G; 101 T; 0 other;	
Alignment Scores:	
Pred. No.: 1.46e-113 Length: 358	
Score: 119.00 Matches: 119	
Percent Similarity: 100.00% Conservative: 0	
Best Local Similarity: 100.00% Mismatches: 0	
Query Match: 25.59% Indels: 0	
DB: 21 Gaps: 0	
US-09-744-313A-1 (1-465) x AAC02755 (1-358)	
QY 216 LeuProAspValAsnLeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGlu 235	
DB 2 CTACCAATGTAAATCTTGGGAAATATATAATCTGTTCTCGAAAACATAATGAAGAG 61	
QY 236 LysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLys 255	
DB 62 AAAGTTCAGCATTTGGAACTTTTATCATCAATTCATTAAATCTTGTGAGTCTCCAAAG 121	
QY 256 ProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLys 275	
DB 122 CCTAAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACTTTCAGAAAACAACAAGAG 181	
QY 276 LeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGln 295	
DB 182 CTTTTCATGATCTGTTTAAATAATATGCAACCGTCTGAAATAACAGAGAGAAACAA 241	
QY 296 AsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyr 315	
DB 242 AATCAGAAATATTATTATGGAGGTGATGATCTGTAGAGAGTCTATGATTACCTGATGAT 301	
QY 316 ValGlyArgValValPheGlnValProAspTyrPheHisLeuLeuMetGlyThr 334	
DB 302 GTAGGACGGGTAGTTTTCAGGTTTCTGAGTGGCTTCATCATCTCTTAAATGGGAACC 358	
RESULT 9	
ABX50426	
ID ABX50426 standard; cDNA; 305 BP.	
XX AC ABX50426;	
XX 25-FEB-2003 (first entry)	
DE Bovine EST associated with lactation/muscle/fat deposition #355.	
XX Bovine; ss; EST; expressed sequence tag; lactation; LMPD;	
KW muscle deposition; fat deposition; genome mapping; gene identification;	
XX gene analysis; cattle breeding.	
XX Bos Taurus.	
XX US2002137160-A1.	

XX PD 26-SEP-2002.  
XX PF 26-OCT-2001; 2001US-0983965.  
XX PR 17-DEC-1998; 98US-113678P.  
XX PR 15-DEC-1999; 99US-0465231.  
XX (BYAT// BYATT J C.  
XX (MATH// MATHIALAGAN N.  
XX (TAON// TAO N.  
XX (WARR// WARREN W C.  
XX  
XX PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX WPI; 2003-102386/09.  
XX  
XX Purified nucleic acid molecules, useful for genome mapping, gene  
PT identification and analysis, cattle breeding or preparation of  
PT constructs for cattle gene expression and genetically improved cattle -  
XX  
XX Claim 2; SEQ ID No 355; 38pp; English.  
XX  
XX The invention relates to a purified nucleic acid molecule associated with  
CC lactation or muscle and fat deposition (designated LMFD), derived  
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a  
CC second nucleic acid molecule comprising any of 5912 nucleotide  
CC sequences, appearing as ABX50072-ABX5393, or complements of them.  
CC Also included are; (1) a transformed cell having a nucleic acid  
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
CC nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridisation between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the complementary  
CC nucleic acid is predictive of the level or pattern of the molecule.  
CC The LMFD nucleic acid is used for determining a level or pattern  
CC of a molecule in a bovine cell or tissue. It is useful for genome  
CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 5912 bovine  
CC LMFD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?DocID=20020137160.  
XX  
XX SQ Sequence 305 BP; 109 A; 60 C; 51 G; 85 T; 0 other;  
Alignment Scores:  
Pred No.: 2,82e-73 Length: 305  
Score: 80.00 Matches: 80  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 17.20% Indels: 0  
DB: 25 Gaps: 0

US-09-744-313a-1 (1-465) x ABX50426 (1-305)

Qy 180 GlnGluTyrLeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuA1a 199  
Db 1 CAGGAATATCTGCAGAAACTCTTCGACATCCAGATTGAGTAACAGTCAACTCTCGCA 60  
Qy 200 AspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysLeuProAspVal 219  
Db 61 GATTTCCTCTCCCAATGGTGGGAAACACAGTTCTTCATAGATACATACAGACGTA 120  
Qy 220 AsnLeuGlyLysIleLeuLysSerValProGlyLysLeuMetLysGlyGlnHis 239

Db 121 AATCTGGGAAATATATAAATCTGTTCTGGAAACTAATGAAGAGAAAGGTCAACAT 180  
Qy 240 LeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSer 259  
Db 181 TTGGAACTTTCATCATGATTTCAATTAATCTTGTGTAATCTCCAAAGCTAAACCGAGT 240  
RESULT 10  
ABX50611  
ID ABX50611 standard; CDNA; 402 BP.  
XX  
XX AC ABX50611;  
XX  
XX DT 25-FEB-2003 (first entry)  
XX DE Bovine EST associated with lactation/muscle/fat deposition #540.  
XX KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
KW muscle deposition; fat deposition; genome mapping; gene identification;  
KW gene analysis; cattle breeding.  
XX  
XX OS Bos Taurus.  
XX  
XX PN US2002137160-A1.  
XX  
XX PD 26-SEP-2002.  
XX  
XX PF 26-OCT-2001; 2001US-0983965.  
XX  
XX PR 17-DEC-1998; 98US-113678P.  
XX PR 15-DEC-1999; 99US-0465231.  
XX  
XX (BYAT// BYATT J C.  
XX (MATH// MATHIALAGAN N.  
XX (TAON// TAO N.  
XX (WARR// WARREN W C.  
XX  
XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX WPI; 2003-102386/09.  
XX  
XX Purified nucleic acid molecules, useful for genome mapping, gene  
PT identification and analysis, cattle breeding or preparation of  
PT constructs for cattle gene expression and genetically improved cattle -  
XX  
XX Claim 2; SEQ ID No 540; 38pp; English.  
XX  
XX The invention relates to a purified nucleic acid molecule associated with  
CC lactation or muscle and fat deposition (designated LMFD), derived  
CC from cattle, and the LMFD nucleic acid can specifically hybridise to a  
CC second nucleic acid molecule comprising any of 5912 nucleotide  
CC sequences, appearing as ABX50072-ABX5393, or complements of them.  
CC Also included are; (1) a transformed cell having a nucleic acid  
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end  
CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
CC nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridisation between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the complementary  
CC nucleic acid is predictive of the level or pattern of the molecule.  
XX The LMFD nucleic acid is used for determining a level or pattern  
XX of a molecule in a bovine cell or tissue. It is useful for genome  
XX mapping, gene identification and analysis, cattle breeding, preparation  
XX of constructs for use in cattle gene expression, or for genetically  
XX improving cattle. The present sequence is one of the 5912 bovine  
XX LMFD EST (expressed sequence tag) nucleic acids.  
XX Note: The present sequence was not shown in the specification but  
XX was obtained in electronic format from the USPTO web site:  
XX seqdata.uspto.gov/sequence.html?DocID=20020137160.

XX SQ Sequence 402 BP; 141 A; 71 C; 76 G; 113 T; 1 other;

Alignment Scores:

Pred. No.: 2,7e-47 Length: 402

Score: 55.00 Matches: 55

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 11.83% Indels: 0

DB: 25 Gaps: 0

US-09-744-313A-1 (1-465) x ABX50611 (1-402)

QY 205 AsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIle 224

Db 3 AATGGTGGGAAACACAGATTCTTGATAGACTACAGACGTAATCTTGGGAAATT 62

QY 225 IleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPhele 244

Db 63 ATAAATCTCTCTCGAAACTTAATCAAGAGAAAGGTCAACATTTCGAACCTTCATC 122

QY 245 MetAsnPheIleAsnSerCysGluSerProLysProLysProSer 259

Db 123 ATGAATTTCATTAATTCTTGTGAATCTCCAAAGCCTAAACCGAGT 167

RESULT 11

ABX77486

ID ABX77486 standard; cDNA; 725 BP.

XX AC ABX77486;

XX XX

DT 12-DEC-2002 (first entry)

XX XX

DE Frog embryonic gene sequence Q9925894.

XX XX

KW Frog; ss; embryonic development; developmental disorder;

KW microarray; cell differentiation.

XX XX

OS Xenopus laevis.

XX XX

PN US2002081610-A1.

XX XX

PD 27-JUN-2002.

XX XX

PF 23-JUL-2001; 2001US-0910943.

XX XX

PR 21-JUL-2000; 2000US-219658P.

XX XX

PA (UVRQ ) UNIV ROCKEFELLER.

XX XX

PI Hemmati-Brivanlou A, Altmann CR;

XX XX

DR WPI; 2002-626534/67.

XX XX

PT Nucleic acid array containing Xenopus embryonic nucleic acids is useful

PT to identify genes involved in embryonic development, to identify

PT different types of embryonic cells, and to diagnose developmental

PT disorders

XX XX

PS Claim 1; Page 756; 823pp; English.

XX XX

CC The invention relates to a nucleic acid array, where each coordinate

CC contains a single nucleic acid species having one of 770 nucleotide

CC sequences (appearing as ABS76747-ABS77516) a of a Xenopus embryonic gene

CC product, or its complement or hybridisable fragment of not less than

CC 20 contiguous nucleotides of one of those sequences. Also included are

CC detecting differential expression of embryonic genes, comprising:

CC (a) contacting a nucleic acid array comprising genes expressed in

CC embryonic but not mature cells with nucleic acids from sample and

CC control cells; and (b) detecting differential hybridisation of nucleic

CC acids from the sample cells relative to the control cells; and detecting

CC defects in development, comprising: (a) contacting nucleic acids from

CC test cells undergoing development with a nucleic acid array of gene

CC products known to play a fundamental role in the development process; and

CC (b) detecting a difference in expression of a fundamental gene in the

CC sample cells relative to a standard. The invention is useful to identify

CC genes involved in embryonic development and related processes such as

CC cell differentiation. This would be useful for diagnosing developmental

CC disorders and for identifying different types of embryonic cells.

XX CC The present sequence is one of the 770 Xenopus embryonic cDNA sequences.

SQ Sequence 725 BP; 229 A; 135 C; 140 G; 218 T; 3 other;

Alignment Scores:

Pred. No.: 1.84e-29 Length: 725

Score: 38.00 Matches: 38

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 8.17% Indels: 0

DB: 24 Gaps: 0

US-09-744-313A-1 (1-465) x ABS77486 (1-725)

QY 243 PheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArgProGlu 262

Db 173 TTTATAATGAATTTCAATAACTCATGTGAATCCCAACCAACCAACAGTAGGCTTGAA 232

QY 263 LeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu 280

Db 233 CTGACTATTCTAGCCGACTTCAGAAAACAATAAAAGCTTTTATGATCTG 286

RESULT 12

AAH69552

ID AAH69552 standard; cDNA; 311 BP.

XX AC AAH69552;

XX XX

DT 19-SEP-2001 (first entry)

XX XX

DE Human cervical cancer marker nucleic acid 826.

XX XX

KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX XX

OS Homo sapiens.

XX XX

PN WO200142467-A2.

XX XX

PD 14-JUN-2001.

XX XX

PF 08-DEC-2000; 2000WO-US33312.

XX XX

PR 08-DEC-1999; 99US-0169681.

XX XX

PR 21-DEC-1999; 99US-0171350.

XX XX

PR 14-MAR-2000; 2000US-0189315.

XX XX

PR 12-MAY-2000; 2000US-0203791.

XX XX

PR 09-JUN-2000; 2000US-0210600.

XX XX

PR 21-JUL-2000; 2000US-0220114.

XX XX

PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX XX

PI Schlegel R, Deeds J, Berger A, Zhao X;

XX XX

DR WPI; 2001-375006/39.

XX XX

PT New isolated nucleic acid for diagnosing and treating cervical cancer

PT and for assessing and detecting compounds for treating the cancer -

XX XX

PS Claim 1; Page 243; 1051pp; English.

XX XX

CC The invention relates to novel genes (AAH69727-AAH73383) associated with

CC cervical cancer with cytostatic activity. The nucleic acids and encoded

CC polypeptides are useful: to assess if a patient is afflicted with

CC cervical cancer or has a pre-malignant condition; to monitor the

CC progression of cervical cancer or a premalignant condition in a patient;

CC and to select and/or assess the efficacy of a compound or therapy for

CC inhibiting cervical cancer in a patient. The nucleic acids may also be

CC useful for gene therapy.  
XX  
SQ Sequence 311 BP; 84 A; 67 C; 63 G; 94 T; 3 other;  
Alignment Scores:  
Pred. No.: 2,32e-16 Length: 311  
Score: 25.00 Matches: 25  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.38% Indels: 0  
DB: 22 Gaps: 0

US-09-744-313A-1 (1-465) x AAH69552 (1-311)  
Qy 430 ValLeuAsnLysGlnLeuThrTyValLeuAspIleValIleGlnGluLeuPhePro 449  
Db 24 GTACTCAACAGCAGCTGACTTATGTTTATTGACATTGTGATACAGGACTGTTTCCA 83  
Qy 450 GluLeuAsnLysVal 454  
Db 84 GAGCTCAATAAGGTA 98

RESULT 13  
AAH70887  
ID AAH70887 standard; cDNA; 313 BP.  
XX  
AC AAH70887;  
XX  
DT 19-SEP-2001 (first entry)  
XX  
DE Human cervical cancer marker nucleic acid 2161.  
XX  
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200142467-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 08-DEC-2000; 2000WO-US33312.  
XX  
PR 08-DEC-1999; 99US-0169681.  
PR 21-DEC-1999; 99US-0171350.  
PR 14-MAR-2000; 2000US-0189315.  
PR 12-MAY-2000; 2000US-0203791.  
PR 09-JUN-2000; 2000US-0210600.  
PR 21-JUL-2000; 2000US-0220114.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Deeds J, Berger A, Zhao X;  
PT WPI; 2001-375006/39.  
DR  
XX  
PT New isolated nucleic acid for diagnosing and treating cervical cancer  
PT and for assessing and detecting compounds for treating the cancer -  
XX  
PS Claim 1; Page 456; 1051pp; English.  
XX  
CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.

US-09-744-313A-1 (1-465) x AAH70887 (1-313)  
Qy 430 ValLeuAsnLysGlnLeuThrTyValLeuAspIleValIleGlnGluLeuPhePro 449  
Db 2 GTACTCAACAGCAGCTGACTTATGTTTATTGACATTGTGATACAGGACTGTTTCCA 61  
Qy 450 GluLeuAsnLysVal 454  
Db 62 GAGCTCAATAAGGTA 76

RESULT 14  
AAH72445  
ID AAH72445 standard; cDNA; 299 BP.  
XX  
AC AAH72445;  
XX  
DT 19-SEP-2001 (first entry)  
XX  
DE Human cervical cancer marker nucleic acid 3719.  
XX  
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200142467-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 08-DEC-2000; 2000WO-US33312.  
XX  
PR 08-DEC-1999; 99US-0169681.  
PR 21-DEC-1999; 99US-0171350.  
PR 14-MAR-2000; 2000US-0189315.  
PR 12-MAY-2000; 2000US-0203791.  
PR 09-JUN-2000; 2000US-0210600.  
PR 21-JUL-2000; 2000US-0220114.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Deeds J, Berger A, Zhao X;  
PT WPI; 2001-375006/39.  
DR  
XX  
PT New isolated nucleic acid for diagnosing and treating cervical cancer  
PT and for assessing and detecting compounds for treating the cancer -  
XX  
PS Claim 1; Page 696; 1051pp; English.  
XX  
CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.

SQ Sequence 299 BP; 82 A; 64 C; 57 G; 96 T; 0 other;  
Alignment Scores:  
Pred. No.: 2,42e-15 Length: 299  
Score: 24.00 Matches: 24  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 5.16% Indels: 0  
DB: 22 Gaps: 0

US-09-744-313A-1 (1-465) x AAH69552 (1-311)  
Qy 430 ValLeuAsnLysGlnLeuThrTyValLeuAspIleValIleGlnGluLeuPhePro 449  
Db 24 GTACTCAACAGCAGCTGACTTATGTTTATTGACATTGTGATACAGGACTGTTTCCA 83  
Qy 450 GluLeuAsnLysVal 454  
Db 84 GAGCTCAATAAGGTA 98

RESULT 13  
AAH70887  
ID AAH70887 standard; cDNA; 313 BP.  
XX  
AC AAH70887;  
XX  
DT 19-SEP-2001 (first entry)  
XX  
DE Human cervical cancer marker nucleic acid 2161.  
XX  
KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200142467-A2.  
XX  
PD 14-JUN-2001.  
XX  
PF 08-DEC-2000; 2000WO-US33312.  
XX  
PR 08-DEC-1999; 99US-0169681.  
PR 21-DEC-1999; 99US-0171350.  
PR 14-MAR-2000; 2000US-0189315.  
PR 12-MAY-2000; 2000US-0203791.  
PR 09-JUN-2000; 2000US-0210600.  
PR 21-JUL-2000; 2000US-0220114.  
XX  
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
PI Schlegel R, Deeds J, Berger A, Zhao X;  
PT WPI; 2001-375006/39.  
DR  
XX  
PT New isolated nucleic acid for diagnosing and treating cervical cancer  
PT and for assessing and detecting compounds for treating the cancer -  
XX  
PS Claim 1; Page 456; 1051pp; English.  
XX  
CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.

SQ Sequence 313 BP; 83 A; 70 C; 63 G; 97 T; 0 other;  
Alignment Scores:  
Pred. No.: 2,33e-16 Length: 313

US-09-744-313A-1 (1-465) x AAH72445 (1-299)

QY 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGlu 450  
Db 2 CTCACACAGCAGCTGACTTATGTTTATTTGGACATTGGATACAGGAAGTGTTCAGAG 61  
QY 451 LeuAsnLysVal 454  
Db 62 CTCATAAGGTA 73

RESULT 15

AAH73007  
ID AAH73007 standard; cDNA; 299 BP.

XX AC AAH73007;

DT 19-SEP-2001 (first entry)

XX Human cervical cancer marker nucleic acid 4281.

DE Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX Homo sapiens.

XX MO200142467-A2.

XX 14-JUN-2001.

PF 08-DEC-2000; 2000WO-US33312.

XX 08-DEC-1999; 99US-0169681.

PR 21-DEC-1999; 99US-0171350.

PR 14-MAR-2000; 2000US-0189315.

PR 12-MAY-2000; 2000US-0203791.

PR 09-JUN-2000; 2000US-0210500.

PR 21-JUL-2000; 2000US-0220114.

XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX Schlegel R, Deeds J, Berger A, Zhao X;

XX WPI; 2001-375006/39.

XX New isolated nucleic acid for diagnosing and treating cervical cancer  
PT and for assessing and detecting compounds for treating the cancer -  
XX Claim 1; Page 941-942; 1051pp; English.  
XX The invention relates to novel genes (AAH68727-AAH73383) associated with  
CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a premalignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.

XX SQ Sequence 299 BP; 82 A; 64 C; 56 G; 95 T; 2 other;

Alignment Scores:

Pred. No.:	2,42e-15	Length:	299
Score:	24.00	Matches:	24
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	5.16%	Indels:	0
DB:	22	Gaps:	0

US-09-744-313A-1 (1-465) x AAH73007 (1-299)

QY 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGlu 450  
Db 2 CTCACACAGCAGCTGACTTATGTTTATTTGGACATTGGATACAGGAAGTGTTCAGAG 61



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 06:24:53 ; Search time 2929 Seconds  
(without alignments)  
3858.512 Million cell updates/sec

Title: US-09-744-313A-1  
Perfect score: 465  
Sequence: 1-MWTFHCFIFRNTQKRGESF.....ELFPELNKVKQKVTSTWSM 465

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 1

Total number of hits satisfying chosen parameters: 45562604

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL-frame\_p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US09744313/runat\_30012004\_113612\_4600/app\_query.fasta\_1.647  
-DB=EST\_QFMT=fastap -SUFFIX=arst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc  
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09744313@cgn\_1\_1\_2810@runat\_30012004\_113612\_4600 -NCFU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

- Database :
- 1: em\_estba.\*
  - 2: em\_esthum.\*
  - 3: em\_estin.\*
  - 4: em\_estmu.\*
  - 5: em\_estov.\*
  - 6: em\_estpl.\*
  - 7: em\_estro.\*
  - 8: em\_hic.\*
  - 9: gb\_est1.\*
  - 10: gb\_est2.\*
  - 11: gb\_hic.\*
  - 12: gb\_est3.\*
  - 13: gb\_est4.\*
  - 14: gb\_est5.\*
  - 15: em\_estfun.\*
  - 16: em\_estom.\*
  - 17: em\_esthum.\*
  - 18: em\_gss\_inv.\*
  - 19: em\_gss\_pln.\*
  - 20: em\_gss\_vrt.\*
  - 21: em\_gss\_fun.\*
  - 22: em\_gss\_mam.\*
  - 23: em\_gss\_mus.\*
  - 24: em\_gss\_pro.\*
  - 25: em\_gss\_rod.\*
  - 26: em\_gss\_phg.\*
  - 27: em\_gss\_vrl.\*
  - 28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	236	50.8	1096	12	BQ053127 AGENCOURT
2	231	49.7	921	13	BQ30190 AGENCOURT
3	220	47.3	664	14	CB136167 K-EST0188
4	208	44.7	643	14	CB114674 K-EST0158
5	206	44.3	1105	10	BG256631 602370809
6	202	43.4	608	12	BM829345 K-EST0102
7	199	42.8	604	14	CB130920 K-EST0180
8	199	42.8	606	12	BM699448 UI-B-DX1-
9	195	41.9	917	13	BM372281 BX372281
10	194	41.7	625	10	BG356529 602564936
11	179	38.5	1076	10	BE874550 601489664
12	176	37.8	817	10	BF892296 602249179
13	170	36.6	513	14	CB142612 K-EST0196
14	168	36.1	506	12	BM835779 K-EST0111
15	167	35.9	773	12	BM982264 UI-CF-EN1
16	166	35.7	548	14	CB139915 K-EST0193
17	164	35.3	494	10	BG547215 602574627
18	164	35.3	726	10	BG432502 602495646
19	162	34.8	638	10	AW957704 EST369774
20	159	34.2	903	9	AL578668 AL578668
21	157	33.8	473	14	CB135556 K-EST0187
22	157	33.8	479	10	AW996088 QV3-BN004
23	157	33.8	570	10	BF104913 601822645
24	157	33.8	818	10	BG191447 RST10541
25	156	33.5	952	13	BQ717610 AGENCOURT
26	152	32.7	497	14	CB137280 K-EST0189
27	152	32.7	558	14	CB160118 K-EST0219
28	151	32.5	457	10	BF978483 602148893
29	150	32.3	717	13	BUS95874 AGENCOURT
30	149	32.0	559	12	BM693800 UI-B-DW1-
31	148	31.8	1043	9	AL574298 AL574298
32	147	31.6	629	13	BUS79763 im90e05.x
33	143	30.8	431	9	AA417301 zu07h02.r
34	140	30.1	428	13	BUS80018 im90e05.y
35	140	30.1	430	10	BE940434 RC3-UT006
36	135	29.0	594	13	BUT38240 UI-B-DW1-
37	135	29.0	640	9	AW327722 dr01d07.x
38	135	29.0	936	10	BG026479 602291871
39	134	28.8	440	9	AW796463 CMI-DM003
40	134	28.8	598	13	BUS83958 UI-CF-ENO
41	132	28.4	421	9	AA431154 zw71f09.r
42	132	28.4	439	13	BQ083624 K-EST0146
43	132	28.4	647	13	BUS79457 UI-CF-DU1
44	130	28.0	986	13	BUI68735 AGENCOURT
45	129	27.7	564	12	BM670092 UI-B-DX1-

ALIGNMENTS

RESULT 1  
BQ053127  
LOCUS AGENCOURT\_6821786 NIH\_MGC\_106 Homo sapiens cdna clone IMAGE:5934711  
DEFINITION 5', mRNA sequence.  
ACCESSION BQ053127  
VERSION BQ053127.1  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Bukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1096)

427	CACGGTTGGTCTCACTCATACATCTCTCAGAGATGCTATATCTGTGAAACACTGAA	486
381	ProArgSerLeuGlnAspLysGlnLysGlyAlaIysGlnThrPheGluGluMetMetAsn	400
487	CCTCGCTCTCTCCAGATATAGCAAAAGGAGCAAAACAGACTTTTGAGAAATGRTGAT	546
401	TyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArg	420
547	TACATTCAGATCTGTTTAGTCAAGTGATTGGTGAGAAACCAAGTATGAAAGCATCAGA	606
421	LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu	440
607	CTTCGTTTGATGCTCTACAGCAACCAAGTACTCAACAGCAGCTGACTTAAGTTTTATTG	666
441	AspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLys	456
667	GACATTGTGATACAGGAACCTGTTTCCAGAGCTCAATAAGGTACAAAG	714

RESULT 2  
BQ230190  
LOCUS  
DEFINITION  
BQ230190 921 bp mRNA  
AGNCOURT 7567151 NIH\_MGC\_70 Homo sapiens cDNA clone IMAGE:6054428  
5', mRNA sequence.

REFERENCES  
1 (bases 1 to 921)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.

found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAMD3313 row: a column: 21  
 High quality sequence stop: 639.

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6054428"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC_70"
/note="Organ: pancreas; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.1 kb. Library constructed by Life Technologies."

```

Alignment Scores:		
Pred. No.:	3,03e-222	Length: 921
Score:	231.00	Matches: 231
Percent Similarity:	100.00%	Conservative: 0
Best Local Similarity:	100.00%	Mismatches: 0
Query Match:	49.68%	Indels: 0
DR:	13	Gaps: 0

QY 221 LeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeu 240

221 LeuGlyLysIleIleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeu 240

NIH-MGC <http://mgc.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
 Tissue Procurement: Dr. Daniel McVicar, DBS/NCI  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)  
 DNA sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ILNL at:  
<http://image.llnl.gov>  
 Plate: LUCM2121 row: e column: 16  
 High quality sequence stop: 675.  
 Location/Qualifiers  
 1. 1096  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5934711"  
 /tissue\_type="natural killer cells, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH MGC 106"  
 /notes="Organ: blood; Vector: pOTB7; Site 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACAGCAG(G). Library constructed by Ling Hong in the  
 laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH MGC Library."  
 355 a 193 c 226 g 305 t 17 others  
 ORIGIN  
 Alignment Scores:  
 Repred. No.: 3.28e-227 Length: 1096  
 Score: 236.00 Matches: 236  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 50.75% Indels: 0  
 DB: 12 Gaps: 0  
 US-09-744-313A-1 (1-465) x BQ053127 (1-1096)  
 QY 221 LeuGlyLysIleLeuLysSerValProGlyLysLeuMetLysGlyLysGlnHisLeu 240  
 Db 7 TTGGGAAATATTAATAATCTGTTCTGGAAAACCTAATGAAAGAGAAAGGTGAGCATTTG 66  
 QY 241 GluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArg 260  
 Db 67 GAACCTTTTATCATGAATTTTCATTAATTTCTGTGAGTCTCCAAAGCGCTAAACCAAGTAGA 126  
 QY 261 ProGlnLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu 280  
 Db 127 CCAGAACTGACCATCTCTCAGCCCTACTTTCAGAAAACAACAGAAAGGTTTTCATATCATCTG 186  
 QY 281 PheLysAsnAlaAsnAtrGAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPhe 300  
 Db 187 TTTTAAATAATGAAACCGTGCTGAAATAACAGAGAGAAAGCAAAATCAGAAATATTTT 246  
 QY 301 MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal 320  
 Db 247 ATGGAGGTGATGACTGTAGAAGGAGTCTATGATTACCTGATGATGATGAGCGGTAGTT 306  
 QY 321 PheGlnValProAspTrrPLeuHisIleLeuMetGlyThrArgIleLeuPheLysAsn 340  
 Db 307 TTTCCAGGTTCTGACTGGCTTTCATCATCTCTTAATGGAACTCGAAATCTCTTTTAAAC 366  
 QY 341 ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGlnLeuPheGlnGlu 360  
 Db 367 ACCCTGGAAATGTACTAGTATCTATCTTCAGTGTAAACTAGAACAGCTATTTCAGGAG 426  
 QY 361 HisArgLeuValSerLeuIleThrLeuArgAspAlaIlePheCysGluAsnThrGlu 380

Db	5	TTGGGAAATATATAATCTGTCCTGGAAACCTAATGAAGAGAAAGTCAGACTTTG	64
Qy	241	GluPropheileMetAsnPhelIeAsnSerCysGluSerProLysProLysProSerArg	260
Db	65	GAACCTTTATCATGAATTCATTAATTCCTGAGTCTCCAAAGCCTAAACCAAGTAGA	124
Qy	261	ProGluLeuThrLleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu	280
Db	125	CCAGAACTGACCATTCCTCAGCCCTACTTCAGAAACCAACAGAGAGCTTTCAATGATCTG	184
Qy	281	PhelYsAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyPhe	300
Db	195	TTTAAATATATGCAACCGTGCTGAAATATACAGAGAGAAAGCAATATCAGATATTTT	244
Qy	301	MetGluValMetThrValGluGlyValTyRaspTyRLeuMetTyRValGlyArgValVal	320
Db	245	ATGGAGGTGATGACTGTAGAAGGACTATGATTAACCTGATGATGTAGGACGGGTAGTT	304
Qy	321	PhelGlnValProAspTyrLeuHisLleLeuMetGlyThrArgLleLeuPheLysAsn	340
Db	305	TTCCAGGTTCTGACTGGCTTCATCATCTCTTAATGGGAACCTCGAATCCTCTTTAAAC	364
Qy	341	ThrLeuGluMetTyThrAspTyRLeuGlnCysLysLeuGluGlnLeuPheGlnGlu	360
Db	365	ACCTCGAAATGATACGATTAATCTTCTAGTGTAACTAGAACAGCTATTTCCAGGAG	424
Qy	361	HisArgLeuValSerLeuLleThrLeuLeuArgAspAlaLlePheCysGluAsnThrGlu	380
Db	425	CACCGTTTGTCTCACTATAACACTTCTCAGAGATGCTATATTTCTGTGAAACACATGAA	484
Qy	381	ProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsn	400
Db	485	CCTCGCTCTCTCAAGATAAGCAAAAGAGGAGCAAAACAGACTTTTGAAGAAATGATGAAT	544
Qy	401	TyrLleProAspLeuValLysCysLleGlyGluThrLysTyRLeuSerLleArg	420
Db	545	TACATTCCAGATCTGTAGTCAAGTGATTTGTGTGAAGAAACCAAGATGAAGCATCAGA	604
Qy	421	LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyRValLeuLeu	440
Db	605	CTTCTGTTGTAGGCTTACAGCAACAGTACTCAACAGCAGCTGACTGTTGTTTATTG	664
Qy	441	AspLleValLleGlnGluLeuPheProGluLeu	451
Db	665	GACATTGTGATACAGAACTGTTCCAGAGCTC	697
RESULT 3			
LOCUS	CB136167		
DEFINITION	K-EST0188577 L14ChoICKO Homo sapiens cDNA clone L14ChoICKO-20-H05		
ACCESSION	CB136167		
VERSION	EST.		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
D AUTHORS	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.		
TITLE	21C Frontier Korean EST Project 2001		
JOURNAL	Unpublished		
COMMENT	Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Eeun-dong Yuseong-gu, Daejeon 305-333, South Korea Tel: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongsung@mail.kribb.re.kr Plate: 20 row: H column: 05		
High quality sequence stop: 664.			
Location/Qualifiers			
1. 664			
/organism="Homo sapiens"			
/mol_type="mRNA"			
/db_xref="taxon:9606"			
/clone="L14ChoICKO-20-H05"			
/sex="M"			
/cell_line="Choi-CK"			
/lab_host="Top10P"			
/clone_lib="L14ChoICKO"			
/note="Organ: Liver; Vector: pTT73-Pac; Site_1: EcoRI; Site_2: NotI; The library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.P., Lennon, G. and Soares, M.B. (1996). Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture."			
BASE COUNT	227 a	107 c	146 g 184 t
ORIGIN			
Alignment Scores:			
Pred. No.:	2,696-211	Length:	664
Score:	220.00	Matches:	220
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	47.31%	Indels:	0
DB:	14	Gaps:	0
US-09-744-313A-1 (1-465) x CB136167 (1-664)			
Qy	14	GlnLysArgGlyGluSerPheGlyLleSerArgLleGlySerLysLysLysGlyValPhe	33
Db	3	CAGAAAGGGGAGAAATCATTGGATTCAGCAAAATAGGTAGCAAAATTAAGAGGTATTC	62
Qy	34	LysSerThrThrMetGluGlyAlaMetLeuProAsnTyRValAlaGluGlyGluAsp	53
Db	63	AAAGTACACATCGAGGAGCTATGTTCCTAATTAATGTGTAGCTCAAGTGAAGAT	122
Qy	54	AspPheLleGluGlyLleValMetGluAspAspSerProValGluAlaValSer	73
Db	123	GATTTATTGAAGAAGTATTGTTGTAATGAAGATGATTCCTCAGTGAGGCTGAGAC	182
Qy	74	ThrProAsnThrProArgAsnLeuAlaTyrLysLleSerLleProTyRValAspPhe	93
Db	183	ACACCTAATACTCCCGAAACCTTGCTGCATGAGAAATAGCAATTCATATGTAGACTTT	242
Qy	94	PheGluAspProSerSerGluArgLysGluLysLysGluArgLleProValPheCysLle	113
Db	243	TTTGAGGATCCCTCCTCTGAAAGAGAGGAGAGAGAGAGAGAGATTCCTGTGTTGATT	302
Qy	114	AspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTyRSerValTyR	133
Db	303	GATGTTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCCTCAACATTGGTCTGTCTAT	362
Qy	134	ArgArgTyRLeuGlnPheTyRValLeuGluSerLysLleThrLlePheHisGlyAlaPhe	153
Db	363	AGAAGATATCTTGAATTTCTATGACTTGAATCAAACTAACAGATTTCAATGTGATTT	422
Qy	154	ProAspAlaGlnLeuProSerLysArgLleLleGlyProLysAsnTyRValPheLeuLys	173
Db	423	CCTGATCCCGAGCTTCTTCTTAAGAGAGATGTTGGCCCCCAAAATTTATGATTTCTAAG	482
Qy	174	SerLysArgGluGlnPheGlnGluTyRLeuGlnLysLeuGlnHisProGluLeuSer	193
Db	483	TCAAGAGGGAGAGTTCCAGAAATATCTACAGAACTTCGAGCATCCAGAACTGAGT	542
Qy	194	AsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAsp	213
Db	543	AATAGTCAACTTCTGGCAGACTTCTTTCCTTAATGGTGGGAGAAACACAAATTTCTGTAT	602
Qy	214	LysLleLeuProAspValAsnLeuGlyLysLleLleLysSerValProGlyLysLeuMet	233
Db	603	AAGTACTACAGATGTAATCTTGGGAAATTTATAAAATCTGTCTCTGGAACTAATG	662

RESULT 4  
CB114674  
LOCUS  
DEFINITION K-EST0158382 L10choCK0 Homo sapiens cDNA clone L10choCK0-1-D02 5',  
mRNA sequence.  
ACCESSION CB114674  
VERSION CB114674.1 GI:27940481  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 1 row: D column: 02  
High quality sequence stop: 643.  
FEATURES  
Location/Qualifiers  
1..643  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="L10choCK0-1-D02"  
/sex="M"  
/cell\_line="Cho-CK"  
/lab\_host="Top10F"  
/notes="Organ: Liver; Vector: pVT3-Pac; Site 1: EcoRI;  
Site 2: NotI; The library was contributed by the Soares  
laboratory and it was constructed as described by Sonaldo,  
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research  
6(9): 791-806. RNA was prepared from harvested cell  
culture."  
BASE COUNT 214 a 123 c 126 g 180 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,33e-199 Length: 643  
Score: 208.00 Matches: 208  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 44.73% Indels: 0  
DB: 14 Gaps: 0  
US-09-744-313A-1 (1-465) x CB114674 (1-643)  
QY 258 ProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPhe 277  
DB 1 CCAAGTCGACGACGACCTCTCTCAGCCCTTCTCAGAAACCAACGAGGCTTTTC 60  
QY 278 AsnAspLeuPheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGln 297  
DB 61 AATGATCTGTTTAAATATATGACACCGTCTGAAATACAGAGAGAAAGCAATACAG 120  
QY 298 AsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGly 317  
DB 121 AATTATTTTATGAGGGTGATGATCTGTAGAGGAGTCTATGATTACGTATGTATGAGA 180  
QY 318 ArgValValPheGlnValProAspTyrPheHisLeuLeuMetGlyThrArgLleLeu 337  
DB 181 CGGGTAGTTTTCCAGGTTCTCTGACTGCTTCATCATCTCTTAATGGGAACCTCGAATCCTC 240

QY 338 PheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeu 357  
DB 241 TTTTAAAAACACCTCGAAATGTATACTGATTAATCTTCACTGTAATAACTAGACAGCTA 300  
QY 358 PheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGlu 377  
DB 301 TTTTCAGGAGCAGCGCTTTGGTCTCACTCAATACACTTCTCAGAGATGCTATATCTGTGAA 360  
QY 378 AsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGlu 397  
DB 361 AACACTGAACCTCGCTCTCTCCAAAGATAAGCAAAAGAGGCAAAACAGACTTTTGAAGAA 420  
QY 398 MetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGlu 417  
DB 421 ATGATGAATPACATTCAGATCTGTGTAGTCAGTGTATTTGGTGAAGAAACCAAGATATGAA 480  
QY 418 SerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyr 437  
DB 481 AGCATCAGACTTCTGTGTGATGCTTACAGCAACAGTACTCAACAAGCAGCTGACTTAT 540  
QY 438 ValLeuLeuAspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGlu 457  
DB 541 GTTTTATGGACATTTGTGATACAGGAACCTGTTTCCAGAGCTCAATAAGGTACAAAGGAA 600  
QY 458 ValThrSerValThrSerTyrMet 465  
DB 601 GTTACCTCTGTGACATCTTGGATG 624  
RESULT 5  
BG256631 1105 bp mRNA linear EST 13-FEB-2001  
LOCUS 602370809F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4478958 5',  
mRNA sequence.  
ACCESSION BG256631  
VERSION BG256631.1 GI:12766447  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 1105)  
NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs@mail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM10310 row: a column: 03  
High quality sequence stop: 645.  
FEATURES  
Location/Qualifiers  
1..1105  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clones="IMAGE:4478958"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_92"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH MGC Library."  
BASE COUNT 380 a 235 c 208 g 282 t  
ORIGIN  
Alignment Scores:



Db 302 CTGAACATGGTCTCTATAGACATATCTTGAATCTATGACTTGAATCAAACTA 361

Qy 147 ThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleleGlyPro 166

Db 362 ACAGAAATTCATGGTGCATTTCTGATGCCAGCTTCCTTCTAAGAGGATCATTTGCCCC 421

Qy 167 LysAsnTyrGluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeu 186

Db 422 AAAAAATATGAATTTCTTAAGTCAAGAGGAGAGAGTTCCTCAAGATATATACAGAACTT 481

Qy 187 LeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGly 206

Db 482 CTGAGCATCCAGAACTAGTAATAGTCAACTTCTGCGAGACTTTCTTCCCTTAATGT 541

Qy 207 GlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleleLys 226

Db 542 GGGGAACACAAATTTCTTGATAGATACACAGATGTAATCTTGGGAATTTATAAA 601

Qy 227 SerVal 228

Db 602 TCTGTT 607

RESULT 7

CB130920 604 bp mRNA linear EST 29-JAN-2003

LOCUS K-EST0180911 L12JSHC0 Homo sapiens cDNA clone L12JSHC0-5-A04 5',

DEFINITION mRNA sequence.

ACCESSION CB130920.1 GI:28095369

VERSION CB130920.1

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 604)

AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

JOURNAL Unpublished

COMMENT Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 5 row: A column: 04

High quality sequence stop: 604.

Location/Qualifiers

1..604

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="L12JSHC0-5-A04"

/sex="M"

/cell\_line="J-SHC"

/lab\_host="Top10P"

/clone\_lib="L12JSHC0"

/note="Organ: Liver; Vector: pCNS-D2; Site 1: EcoRI; Site 2: NotI; The poly (A) + RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoRI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10P by electroporation method.

The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT 205 a 100 c 137 g 162 t

ORIGIN

Alignment Scores:

Pred. No.: 3,76e-190 Length: 604

Score: 199.00 Matches: 199

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 42.80% Indels: 0

DB: 14 Gaps: 0

US-09-744-313A-1 (1-465) x CB130920 (1-604)

Qy 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyLysSerArgIleleGlySerLysIleLys 30

Db 7 AGGAACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGATAGGTAGCAAAATATAA 66

Qy 31 GlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGlu 50

Db 67 GGAGTATTCAAAAGTACCACAAATGGAGGAGCTATGTTCCTTAATATGTTAGTCTGAA 126

Qy 51 GlyGluAspAspPheIleGluGlyValValMetGluAspAspSerProValGlu 70

Db 127 GGTGAAGATGATTTTATTTAGAGAGGATTTGTTGTAATGAGATGATTTCTCCAGTGAG 186

Qy 71 AlaValSerThrProAsnThrProArgAsnLeuAlaLaTrpLysIleSerIleProTyr 90

Db 187 GCTGTGAGCACACTAATACTCCCGAAACCTTGCTGCATGCGAAATATAGCATTCATAT 246

Qy 91 ValAspPhePheGluAspProSerSerGluArgLysGluLysLysGluArgIleProVal 110

Db 247 CTAGACTTTTTTGGAGTATCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGG 306

Qy 111 PheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrp 130

Db 307 TTTTGTATGTATGTGTAAGAGAAATGATAGAGACAGTGTGACACAGAGCTGACATGG 366

Qy 131 SerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150

Db 367 TCTGTCTATAGAGATATCTTGAATTTCTATGACTTCAATCAAACTAAACAGATTTCA 426

Qy 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleleGlyProLysAsnTyrGlu 170

Db 427 GGTGCATTTCTGATGCCAGCTTCTTCTAAGAGGATCATTTGGCCCCCAAAATATGAA 486

Qy 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHisPro 190

Db 487 TTTTAAAGTCAAGAGGAGGAGGAGTTCCAAGAAATATCTACAGAACTTCTGCGAGATCCA 546

Qy 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209

Db 547 GAACGTGAGTAATAGTCAACTTCTGCGAGACTTTCTTTCCTTATGTTGGGGAACA 603

RESULT 8

BM699448 606 bp mRNA linear EST 28-FEB-2002

LOCUS UI-E-DX1-agg-g-17-0-UI-r1 UI-E-DX1 Homo sapiens cDNA clone

DEFINITION UI-E-DX1-agg-g-17-0-UI 5', mRNA sequence.

ACCESSION BM699448

VERSION BM699448

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 606)

AUTHORS Bonaldo,M.P., Lennon,G. and Soares,M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)

MEDLINE 97044477



889548  
Contact: Soares, MB  
Coordinated Laboratory for Computational Genomics  
University of Iowa  
375 Newton Road, 4156 MEHRP, Iowa City, IA 52242, USA  
Tel: 319 335 8250  
Fax: 319 335 9565  
Email: bento-soares@uiowa.edu  
Tissue Procurement: Dr. Gregg Hageman  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).  
Seq primer: M13 Reverse.  
Location/Qualifiers  
1. .606  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-DX1-agv-g-17-0-UI"  
/tissue\_type="fetal eyes"  
/dev\_stage="fetal"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-DX1"  
/notes="Organ: eye; Vector: p773-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoR I; Site 2: Not I;  
UI-E-DX1 is a normalized cDNA library containing the  
following tissue(s): fetal eyes. The library was  
constructed according to Bonaldo, Lennon and Soares,  
Genome Research, 6:791-806, 1996. First strand cDNA  
synthesis was primed with an oligo-dT primer containing a  
Not I site. Double stranded cDNA was ligated to an EcoR I  
adaptor, digested with Not I, and cloned directionally  
into p773-Pac vector. The oligonucleotide used to prime  
the synthesis of first-strand cDNA contains a library tag  
sequence that is located between the Not I site and the  
(dT)18 tail. The sequence tag for this library is  
AGAAATCAAGA. This library was created for the program, Gene  
Discovery in the Visual System, supported by National Eye  
Institute (NEI)."  
204 a 114 c 118 g 170 t  
BASE COUNT  
ORIGIN  
Alignment Scores:  
Pred. No.: 3,77e-190 Length: 606  
Score: 199.00 Matches: 199  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 42.80% Indels: 0  
DB: 12 Gaps: 0  
US-09-744-313A-1 (1-465) x BM699448 (1-606)  
QY 267 SerProThrSerGluAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsn 286  
DB 2 AGCCTACTTCAGAAACACACAGAGCTTTTCATGATCTGTTAAATATATGCAAC 61  
QY 287 ArgAlaGluAsnThrGluArgLysGluAsnGlnAsnTyrPheMetGluValMetThrVal 306  
DB 62 CGTGCTGAATATACAGAGAGAAACAAATCAGAAATATTATTATGGAGGTGATGACTGTA 121  
QY 307 GluglyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyr 326  
DB 122 GAAGGAGTCTATGATTAATCTGATGATGATGAGAGCGGTAGTTTCCAGGTTCCTGACTGG 181  
QY 327 LeuHisHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThr 346  
DB 182 CTTTCATCATCTCTTAATGGGACATCGAATCTCTTTTAAACACCCCTGGAAATGATPACT 241  
QY 347 AspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeu 366  
DB 242 GATTACTATCTTCAGTAACTAGACAGCTATTTCAGGAGCACCGTTGGTCTCACTC 301

QY 367 IleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAsp 386  
DB 302 ATACACTTCTCAGAGATGCTATATTTCTGTGAACAACTGAACTGCTCTCTCCAGAT 361  
QY 387 LysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeu 406  
DB 362 AAGCAAAAGAGCAAAACAGACTTTTGAAGAATGATGATGATGATGATGATGATGAT 421  
QY 407 ValLysCysIleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeu 426  
DB 422 GTCAAGTGATTTGGTGAAGAAACCAAGATATCAAGATCAGACTTCTGTTGATGGCTTA 481  
QY 427 GlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGlu 446  
DB 482 CAGCAACAGTACTCAACAGCAGCTGACTATGTTTATGGACATGTGATGATGATGAT 541  
QY 447 LeuPheProGluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
DB 542 CTGTTTCAGAGCTCAATATAGTACAAAGAGTACCTCTCTGATGATGATGATGATG 598  
RESULT 9  
BX372281/c  
LOCUS BX372281 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED  
DEFINITION Homo sapiens cDNA clone CS0DL007YB10 3-PRIME, mRNA sequence.  
ACCESSION BX372281  
VERSION BX372281.1 GI:30441965  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 917)  
AUTHORS Li W.B., Gruber C., Jesses, J. and Polayes D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr  
Library was constructed by Life Technologies, a division of  
Invitrogen. This sequence belongs to sequence cluster 3600.r For  
more information about this cluster, see  
http://www.genoscope.cns.fr/  
cgi-bin/cluster.cgi?seq=CS0BAI0442B11\_CS04210\_1&cluster=3600.r.  
Contact : Peng Liang Email : fliang@lifetech.com URL :  
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600  
Faraday Avenue Genoscope sequence ID : CS0BAI0442B11\_CS04210\_1.  
Location/Qualifiers  
1. .917  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CS0DL007YB10"  
/cell\_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"  
/cell\_lines="RAMOS CELL LINE"  
/clone\_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT  
25-NORMALIZED"  
/note="1st strand cDNA was primed with a NotI-oligo(dT)  
primer. Five prime end enriched, double-strand cDNA was  
digested with Not I and cloned into the Not I and EcoR V  
sites of the pCMVSPORT 6 vector. Library was normalized."  
BASE COUNT 257 a 168 c 177 g 310 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 6,38e-186 Length: 917  
Score: 195.00 Matches: 253  
Percent Similarity: 98.83% Conservative: 0  
Best Local Similarity: 98.83% Mismatches: 1  
Query Match: 41.94% Indels: 3  
DB: 13 Gaps: 0

US-09-744-313A-1 (1-465) x BX372281 (1-917)

QY 181 GluTyrLeuGlnValLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAsp 200  
 Db 777 GAATATCTACAGAACTCTGCGAGTCTCAGACTGAGTATAGTCACTTCTGGCAGAC 718  
 QY 201 PheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsn 220  
 Db 717 TTTCTTCCCTAATGGTGGGAAACACAATTTCTTGATAGATACTACCAGATGTTAAT 658  
 QY 221 LeuGlyLysIleLys-SerValProGlyLysLeuMetLysGlyGlnHisLe 240  
 Db 657 CTTGGGAAATATATANA-ATCTGTCTCTGGAAACTAATGAAGAGAAAGCTCAGCATTT 599  
 QY 240 uGluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProSerAr 260  
 Db 598 GGAACCTTTTATCATGAATTTTCAATTAATCTTGAGTCTCCAAAGCTTAAACCAAGTAG 539  
 QY 260 ProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLe 280  
 Db 538 ACAGACTGACATCTCTCAGCCCTACTCTCAGAAACACACAGAGAGCTTTTCAATGATCT 479  
 QY 280 uPheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPh 300  
 Db 478 GTTTAAATAATATGCAACCGTCTGAAATATACAGAGAGAAAGCAAAATCAGATTTATT 419  
 QY 300 eMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVa 320  
 Db 418 TATGGAGGTGATGACTGTAGAAGAGTCTATGATTAACCTGATGTATGAGACGGGTAGT 359  
 QY 320 lPheGlnValProAspTyrLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAs 340  
 Db 358 TTTCCAGGTTCTGAGTGGCTTCATCATCTTAATGGAACTCGATCTCTTTAAAAA 299  
 QY 340 nThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGl 360  
 Db 298 CACCTCGGAAATGTAATGATTAATCTTCACTGTTAACTAGAACAGCTATTTTCAGGA 239  
 QY 360 uHisArgLeuValSerLeuThrLeuLeuHisLeuMetGlyThrArgIleLeuPheLysAs 380  
 Db 238 GCACCGTTTGGTCTCACTCATACACTTCTCAGAGATGCTATATTTCTGTGAAACACTGA 179  
 QY 380 uProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAs 400  
 Db 178 ACCTCGCTCTCTCCAGATAAGCAAAAGAGCAAAACAGACTTTTGAAGAAATGATGAA 119  
 QY 400 nTyrIleProAspLeuValLysCysIleGlyGluGluThr-LysTyrGluSerIleA 420  
 Db 118 TTACATTCAGATCTGTAGTCAAGTGTATTGTTGAGAAACNCACAGATATGAAGCATCA 59  
 QY 420 rgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGln 434  
 Db 58 GACTTCTGTTTGTGGCTTACAGCAACCACTACTCAACAGCAG 15

RESULT 10  
 BG536529 625 bp mRNA linear EST 03-APR-2001  
 LOCUS 602564936F1 NIH\_MGC\_77 Homo sapiens cDNA clone IMAGE:4689464 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BG536529  
 VERSION BG536529.1 GI:13528075  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 625)  
 AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1CM1505 row: h column: 09  
 High quality sequence stop: 625.

#### FEATURES

source

1. 625  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4689464"  
 /lab\_host="DH10B (T1 phage-resistant)"  
 /clone\_lib="NIH MGC 77"  
 /note="Organ: lung; Vector: pDNR-LIB (Clontech); Site 1:  
 SfiI (ggccgctcgcc); Site 2: SfiI (ggccattatggc); 5' and  
 3' adaptors were used in cloning as follows: 5' adaptor  
 sequence: 5'-CACGCCATTATGCC-3' and 3' adaptor sequence:  
 5'-ATTCTAGAGCGGCGGCGGACATG-dt(30)BN-3' (where B = A,  
 C, or G and N = A, C, G, or T). Average insert size 1.9  
 kb (range 0.5-4.0 kb). 12/15 colonies contained inserts  
 by PCR. This library was enriched for full-length clones  
 and was constructed by Clontech Laboratories (Palo Alto,  
 CA). Note: this is a NIH MGC Library."

BASE COUNT 211 a 123 c 119 g 172 t

#### ORIGIN

Alignment Scores:  
 Pred. No.: 4,33e-185 Length: 625  
 Score: 194.00 Matches: 194  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 41.72% Indels: 0  
 DB: 10 Gaps: 0  
 US-09-744-313A-1 (1-465) x BG536529 (1-625)

QY 248 lLeAsnSerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSer 267  
 Db 2 ATTAATTTCTGTGAGTCTCCAAAGCTTAAACCAAGTAGACCAACTGACCATCTCAGC 61  
 QY 268 ProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArg 287  
 Db 62 CTTACTTCAGAAACACACAGAGAGCTTTTCATGATCTGTTTAAATAATATGCAACCGT 121  
 QY 288 AlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGlu 307  
 Db 122 GCTGAAATACAGAGAGAAAGCAAAATCAGAAATTAATTTATGAGAGGTGATGCTAGTAA 181  
 QY 308 GlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeu 327  
 Db 182 CGAGTCTATGATTAACCTGATGATGTATGAGGACGGGTAGTTTTCAGGGTTCCTGCTGCTT 241  
 QY 328 HisHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAsp 347  
 Db 242 CATCATCTCTTAATGGGAACCTCGAATCCTCTTTAAACACACCTCGAATGTAATGATCAT 301  
 QY 348 TyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGlnHisArgLeuValSerLeuIle 367  
 Db 302 TACTATCTTCAGTGTAACTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTCACTCATA 361  
 QY 368 ThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLys 387  
 Db 362 ACACCTTCTCAGAGATGCTATATTTCTGTGAAACACTCAACCTCGCTCTCTCCAGATGAG 421  
 QY 398 GlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuVal 407  
 Db 422 CAAAAGAGGCAAAACAGACTTTTGAAGAAATGATGAATTACATTCCAGATCTCTTAGTC 481  
 QY 408 LysCysIleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGln 427

Tissue Procurement: CLONTECH Laboratories, Inc.  
 cDNA Library Preparation: CLONTECH Laboratories, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>

Plate: L1CM1505 row: h column: 09  
 High quality sequence stop: 625.



FEATURES		Location/Qualifiers	
source		1.. 513	
		/organism="Homo sapiens"	
		/mol_type="mRNA"	
		/db_xref="taxon:9606"	
		/clone="L16HLK3-25-H03"	
		/cell_line="HLK-3"	
		/lab_host="Top10F"	
		/clone_lib="L16HLK3"	
		/note="Organ: Liver; Vector: pTV73-Pac; Site 1: EcoRI; Site 2: NotI; The library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), Genome Research 6(9): 791-806. RNA was prepared from harvested cell culture."	
BASE COUNT	165 a	89 c	112 g 147 t
ORIGIN			
Alignment Scores:			
Pred. No.:	5.78e-161	Length:	513
Score:	170.00	Matches:	170
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	36.56%	Indels:	0
DB:	14	Gaps:	0
US-09-744-313A-1 (1-465) x CB142612 (1-513)			
QY	34	LysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGluGlyGluAsp	53
Db	3	AAAAGTACCACAATGGAGGAGCTATGTTGCCCTAAATTATGGTGTAGCTGAAGGTGAAGAT	62
QY	54	AspPheIleGluGlyIleValValMetGluAspAspSerProValGluAlaValSer	73
Db	63	GAITTTATTGAAGAGGTATTGTTGTAATGGAAGATGATCTCCAGTGGAGGCTGTGAGC	122
QY	74	ThrProAsnThrProArgAsnLeuAlaIleTyrLysIleSerIleProTyrValAspPhe	93
Db	123	ACACCTAATACTCCCCGAAACCTTGCTGCATCGAAAAATTAGCATTCCATATGTAGACTTT	182
QY	94	PheGluAspProSerSerGluArgLysGlyLysGluArgIleProValPheCysIle	113
Db	183	TTTGAGGATCCCTCTCTGAAAGGAGGAGAGAAAAAGAAATTCCTGTGTTTGTATT	242
QY	114	AspValGluArgAsnAspArgArgAlaValGlyHisGluProGluHisTyrSerValTyr	133
Db	243	GATGTTGAAAGAAATGATGAAGAGCAGTGGACAGCCTGAACACTTGGTCTGTCTAT	302
QY	134	ArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAlaPhe	153
Db	303	AGAGATATCTTGAATTTCTATGTTGATCAAACTAACAGAAATTCATGTTGTCATTT	362
QY	154	ProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGluPheLeuLys	173
Db	363	CCTGATGCCAGCTTCCTTCTTAAGAGGATCATTTGGCCCCCAAAAAATATGATTTCTTAAG	422
QY	174	SerLysArgGluGluPheGlnGlyTyrLeuGlnLysLeuLeuGlnHisProGluLeuSer	193
Db	423	TCAAAGAGGGAAGAGTTCCAAGATATCTACAGAACTTCTGACAGAACTTCTGACAGACTCCAGAACTGAGT	482
QY	194	AsnSerGlnLeuLeuAlaAspPheLeuSer	203
Db	483	AATAGTCAACTTCTGGCAGACTTTCTTTCC	512
RESULT 14			
LOCUS	BM835779	506 bp	mRNA linear EST 06-MAR-2002
DEFINITION	K-EST0111014 S11SN1 Homo sapiens cDNA clone S11SN1-73-D02 5', mRNA sequence.		
ACCESSION	BM835779		
VERSION	BM835779.1		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		

Alignment Scores:		8.38e-167	Length: 817
Pred. No.:	176.00	Matches: 176	
Score:	100.00%	Conservative: 0	
Percent Similarity:	100.00%	Mismatches: 0	
Best Local Similarity:	100.00%	Indels: 0	
Query Match:	37.85%	Gaps: 0	
DB:	10		
US-09-744-313A-1 (1-465) x BF692296 (1-817)			
QY	253	SerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsn	272
Db	2	TCTCCAAAGCCTAAACCAAGTAGACCAAGACTGACCATCTCAGCCCTACTTCAGAAAAC	61
QY	273	AsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGlu	292
Db	62	AACAAGAAGCTTTCAATGATCTGTTTAAATAATATGCAACCGCTGCAAAATACAGAG	121
QY	293	ArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyr	312
Db	122	AGAAACCAAAATCAGAAATTAATTTATGGAGGTGATGACTGTAGAGGAGTCTATGATTAC	181
QY	313	LeuMetTyrValGlyArgValPheGlnValProAspTyrLeuHisLeuLeuMet	332
Db	182	CTGATGTATGAGGAGCGGTAGTTTCCAGGTTCTCTGACTGGCTTCATCATCTCTTAATG	241
QY	333	GlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCys	352
Db	242	GGAACTCGAATCCTCTTTAAACACCCCTGGAATATGATATCTACTATCTTCAGTGT	301
QY	353	LysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAsp	372
Db	302	AAACTAGACAGCTATTTTCAGGAGCACCGTTTGCTCTCACTCATACACTTCTCAGAGAT	361
QY	373	AlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLys	392
Db	362	GCTATATCTGTGAACACTGAACCTCGCTCTCTCCAGATAGCAAAAGAGCAAAA	421
QY	393	GlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGlu	412
Db	422	CAGACTTTTGAAGAAATGATGAATATCAATTCAGATCTGTAGTCAAGTGTATGGTGA	481
QY	413	GluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGln	428
Db	482	GAAACCAAGATGAAAGCATCAGACTTCTGTGTTGAGCTTACAGCAA	529
RESULT 13			
LOCUS	CB142612	513 bp	mRNA linear EST 29-JAN-2003
DEFINITION	K-EST0196459 L16HLK3 Homo sapiens cDNA clone L16HLK3-25-H03 5', mRNA sequence.		
ACCESSION	CB142612		
VERSION	1 (bases 1 to 513)		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.B., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.		
TITLE	21C Frontier Korean EST Project 2001		
JOURNAL	Unpublished		
COMMENT	Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea Tel: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongseung@mail.kribb.re.kr Plate: 25 row: H column: 03 High quality sequence stop: 513.		

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 506)  
AUTHORS Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,  
Oh, K.J., Cheong, J.S., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and  
Kim, Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished  
COMMENT Genom Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongseung@mail.kribb.re.kr  
Plates: 73 row: D column: 02  
High quality sequence stop: 506.  
Location/Qualifiers  
1. .506  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S11SNU1-73-D02"  
/sex="M"  
/tissue type="Stomach"  
/cell\_type="Lymphoblast-like"  
/cell\_line="SNU-1"  
/lab\_host="Top10r"  
/clone\_lib="S11SNU1"  
/notes="Organ: Stomach; Vector: pME18-FL3; Site: 1: XhoI;  
Site: 2: XhoI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tobacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including SfiI  
site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized with Superscript II using SfiI  
oligo-dT primer. After first strand synthesis, RNA was  
degraded by NaOH treatment and cDNA was amplified by PCR  
reaction. The PCR products were digested with SfiI and  
cloned into DraIII- digested pME18-FL3 vector. The  
obtained cDNA vectors were used for transfection of  
competent cells E. coli Top10r by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."  
BASE COUNT 158 a 94 c 104 g 150 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 5,94e-159 Length: 506  
Score: 168.00 Matches: 168  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 36.13% Indels: 0  
DB: 12 Gaps: 0  
US-09-744-313A-1 (1-465) x BM835779 (1-506)  
Qy 298 AsnThrGluProArgSerLeuGlnAspLysGlnLysGlnLysGlnThrPheGluGlu 317  
Db 2 AATATTTTATGGAGGTGATGACTGTGAGAGGAGTCTATGATACCTGATGATGAGGA 61  
Qy 318 ArgValValPheGlnValProAspTrpLeuHisLeuLeuMetGlyThrArgIleLeu 337  
Db 62 CGGTAGTTTCCAGGTTCTGACTGGCTTCATCATCTCTTAATGGGAACCTGAATCCTC 121  
Qy 338 PheLysAsnThrLeuGluMetTyThrAspTyThrLeuGlnCysLysLeuGlnLeu 357  
Db 122 TTTAAACACCCCTGGAATGATATCTGATCTATCTTTCAGTGTAAACAGAGCTA 181  
Qy 358 PheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGlu 377  
Db 182 TTTCAGGAGCACCGTTGGTCTCTACTATACACTTCTCAGAGATGCTATATCTGTGAA 241

378 AsnThrGluProArgSerLeuGlnAspLysGlnLysGlnLysGlnThrPheGluGlu 397  
Db 242 AACACTGACCTCGCTCTCTCCAGATAGCAAAAGGAGCAAAACAGACTTTTGAAGAA 301  
Qy 398 MetMetAsnTyrlleProAspLeuValLysCysLleGlyGluGluThrLysTyGlu 417  
Db 302 ATGATGAATTACATTCAGATCTGTGTAGTCAGTGTATTGGTGAAGAAACCAAGTATGAA 361  
Qy 418 SerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTy 437  
Db 362 AGCATCAGACTTCTGTTGATGGCTTACAGCAACCACTACTCAACAGCAGCTGACTTAT 421  
Qy 438 ValLeuLeuAspLysValIleGlnGluLeuPheProGlnLeuAsnLysValGlnLysGlu 457  
Db 422 GTTTTATTGGACATTTGTGATACAGGAACCTGTTCCAGAGCTCAATAGGTACAAAGGAA 481  
Qy 458 ValThrSerValThrSerTrpMet 465  
Db 482 GTTACCTCTGTGACATCTTGGATG 505  
RESULT 15  
BM982264/c  
LOCUS BM982264  
DEFINITION UI-CF-EN1-acr-j-04-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone  
UI-CF-EN1-acr-j-04-0-UI 3', mRNA sequence.  
ACCESSION BM982264  
VERSION BM982264.1 GI:19605588  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 773)  
AUTHORS Bonaldo, M.F., Lennon, G. and Soares, M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery  
JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
PubMed 889548  
COMMENT Contact: McCray, PB  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com) or from Open Biosystems  
(www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA-Yes.  
Location/Qualifiers  
1. .773  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-EN1-acr-j-04-0-UI"  
/tissue\_type="Primary Lung Cystic Fibrosis Epithelial  
Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-EN1"  
/note="Organ: Lung; Vector: p773-Pac (Pharmacia) with a  
modified polylinker; Site: 1: EcoR I; Site 2: Not I;  
UI-CF-EN1 is a normalized cDNA library containing the  
following tissue(s): Primary Lung Cystic Fibrosis  
Epithelial Cells. The library was constructed according to  
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,

1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.

TAG LIB=UI-CF-ENI  
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS  
6hr to LPS 24h  
TAG SEQ=CTGCTCAGGT"

BASE COUNT 233 a 146 c 124 g 269 t 1 others  
ORIGIN

Alignment Scores:  
Pred. No.: 9,53e-158 Length: 773  
Score: 167.00 Matches: 183  
Percent Similarity: 99.46% Conservative: 0  
Best Local Similarity: 99.46% Mismatches: 1  
Query Match: 35.91% Indels: 1  
DB: 12 Gaps: 0

US-09-744-313A-1 (1-455) x BM982264 (1-773)

QY	282	LYSASNALASNAARGALAGLUASNTHRGIUARGLYSGINASGLNASNTYRPHMET	301
DB	740	AAAAATATGCAACCGTCTGAAATACAGAGAGAAAGCAAAATCAGAA-TATTTTATG	682
QY	302	GLUVALMETHRVALGLUGLYVALTYRASP TYRLEUWETTYRVALGLYARGVALVALPHE	321
DB	681	GAGGTGATGACTGTAGAGGAGTCTATGATTACCTGATGTATGTAGGACGGGTAGTTTC	622
QY	322	GLNVALPROASPTPLLEUHIHISLEUWETGLYTHRARGILEUPEHLYSASNTHR	341
DB	621	CAGGTCTCTGACTGGCTTCATCATCTCTTAATGGGAACTCGAATCTCTTTAAAACACC	562
QY	342	LEUGLUMETTYRTHRASP TYR TYRLEUGLINCYSLYSLEUGLULNLEUPHEGLNGLHIS	361
DB	561	CTGGAAATGTATCTACTGATTAATCTTCAGTGTAACTAGAACAGCTATTTTCAGAGCAC	502
QY	362	ARGLEUVALSERLEULETHRLEUARGASPALAILPEHCYSGLUASNTHRGLUPRO	381
DB	501	CGTTTGGTCTCACTATACACTTCTCAGAGATGTATATTTCTGTGAAAACACCTGAACCT	442
QY	382	ARGSERLEUGLNASPLYSGLINLYSGLYALALYSGLNTHRPHGLUGLUMETMETASNTYR	401
DB	441	CGCTCTCTCCAGATAGCAAAAAGGAGCAAAAACAGACTTTTGAGAAATGATGAATTAC	382
QY	402	ILEPROASPLEUVALLYSCYSILEGLYGLULUTHRLYSTYRGLUSERILEARGLEU	421
DB	381	ATTCCAGATCTGTGTAGTCAAGTGTATTTGGTGAGAACCAAGTATGAAAGCATCAGACTT	322
QY	422	LEUPHEASPGLYLEUGLINPROVALLEUASNLYSGLINLEUTHRTYRVALLEULEUASP	441
DB	321	CTGTTTGTATGCTTACAGCAACCACTACTCAACAAGCAGCTGACTTATGTTTATTGGAC	262
QY	442	ILEVALLEGLNGLUPEHLEUPROGLUASNTYRVALGLINLYSGLUVALTHRSEVAL	461
DB	261	ATTGTGATACAGGAACGTGTTTCCAGAGCTCAATAAGGTACAAAAGGAAGTTACCTCTGTG	202
QY	462	THRSERTPMET	465
DB	201	ACATCTTGGATG	190

Search completed: January 31, 2004, 09:03:14  
Job time : 2945 secs



and is derived by analysis of the total score distribution.

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 02:37:46 ; Search time 377 Seconds  
(without alignments)  
3329.546 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 2437

Sequence: 1 MYLIHCLIFRNQKRGESF.....ELPPELNKVKQKVTGVTSMW 465

Scoring table

BLOSUM62				
Xgapop 10.0	Xgapext 0.5			
Ygapop 10.0	Ygapext 0.5			
Fgapop 6.0	Fgapext 7.0			
Delop 6.0	Delext 7.0			

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-Q=/cgn2 1/USPTO.spool/US09744313/runat 30012004 113535 4176/app\_query.fasta.1.647  
-DB=N Geneseq 19Jun03 -OPMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313 @CNG 1.0 @runat 30012004 113535 4176 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORE=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOB=6  
-FGAPEXT=7 -YGAPOB=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

Database : N Geneseq 19Jun03:\*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:\*  
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:\*  
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:\*  
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:\*  
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:\*  
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:\*  
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:\*  
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:\*  
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:\*  
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:\*  
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:\*  
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:\*  
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:\*  
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:\*  
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:\*  
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:\*  
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:\*  
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:\*  
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:\*  
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:\*  
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:\*  
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:\*  
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:\*  
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:\*  
25: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2437	100.0	1992	22	SNEXN DNA #1. Hom
2	2386	97.9	1716	22	Human polynucleoti
3	2386	97.9	2176	22	Human polynucleoti
4	2386	97.9	2175	22	Human polynucleoti
5	2382	97.7	3145	24	CDNA encoding a pr
6	2057.5	84.4	1551	22	Human polynucleoti
7	1054	43.2	779	20	Human gene express
8	981.5	40.3	725	24	Frog embryonic gen
9	682	28.0	402	25	Bovine EST associa
10	634	26.0	358	21	Human secreted pro
11	505	20.7	305	25	Bovine EST associa
12	262	10.8	1900	24	CDNA encoding huma
13	255	10.5	2133	22	Human polynucleoti
14	255	10.5	2181	22	Human polynucleoti
15	249.5	10.2	2494	24	CDNA sequence #238
16	235	9.6	2456	22	CDNA encoding nove
17	228	9.4	2869	23	DNA encoding novel
18	192	7.9	1429	24	Human secreted pro
19	192	7.9	1429	24	CDNA encoding huma
20	188	7.7	2955	22	Human cDNA sequenc
21	184.5	7.6	3585	23	DNA encoding novel
22	182.5	7.5	3435	23	Drosophila melanog
23	156	6.4	2310	24	Hypoxia-regulated
24	156	6.4	4156	21	Human ORFX ORF3067
25	155.5	6.4	2404	21	Human serum and gl
26	155.5	6.4	2483	22	Human full-length
27	155.5	6.4	2572	23	DNA encoding novel
28	155.5	6.4	2702	23	DNA encoding novel
29	155.5	6.4	2711	22	Human polynucleoti
30	155.5	6.4	2760	22	Human polynucleoti
31	155	6.4	2250	22	Novel protein kina
32	152.5	6.3	1294	23	CDNA encoding nove
33	152.5	6.3	1294	23	DNA encoding novel
34	149.5	6.1	2512	20	Human serum glucoc
35	148.5	6.1	1607	21	Human ORFX ORF2906
36	148	6.1	816	25	Aspergillus oryzae
37	147.5	6.1	1213	24	Human intracellular
38	147.5	6.1	1502	24	Human cDNA encodin
39	147.5	6.1	3106	23	Human prostate exp
40	146.5	6.0	810	22	Human cDNA 5'-end
41	146.5	6.0	810	22	Human cDNA clone r
42	142.5	5.8	311	22	Human cervical can
43	142.5	5.8	313	22	Human cervical can
44	142	5.8	995	24	Human secreted pro
45	141.5	5.8	1940	21	Human secreted pro

## ALIGNMENTS

RESULT 1  
AAC86397  
ID AAC86397 standard; DNA; 1992 BP.  
XX  
AC AAC86397;  
XX  
XX 21-MAR-2001 (first entry)  
DT  
XX  
XX SNEXN DNA #1.  
DE  
XX SNEXN; human; sorting nexin; inflammation; asthma; allergy; AIDS;  
KW neurological disorder; gastrointestinal; smooth muscle cell;  
KW cancer; gene therapy; ds.  
XX  
XX Homo sapiens.  
XX  
XX WO200073334-A2.  
PN

XX	07-DEC-2000.
PD	
XX	26-MAY-2000; 200OWO-US14831.
XX	
PR	27-MAY-1999; 99US-01136740.
PR	16-JUN-1999; 99US-0139566.
XX	(INCY-) INCYTE GENOMICS INC.
PA	
XX	Yue H, Tang YT, Azimzai Y;
XX	WPI; 2001-041141/05.
DR	
XX	Novel Sorting Nexin polypeptides and polymucleotides useful for
PT	diagnosing and treating disorders associated with their expression e.g.
PT	autoimmune disorders, smooth muscle cell disorders and cell
PT	proliferative disorders -
XX	Claim 5; Page 82-83; 84pp; English.
PS	
CC	The present invention relates to human sorting nexin (SNEYN).
CC	Compositions containing SNEYN or agonists of SNEYN are useful for
CC	treating a disease or condition associated with decreased expression of
CC	functional SNEYN and compositions containing antagonists of SNEYN are
CC	useful for treating a disease or condition associated with
CC	overexpression of functional SNEYN. These can be immune
CC	disorders for example inflammation, asthma, allergy, and AIDS,
CC	neurological disorders, gastrointestinal disorders, smooth muscle cell
CC	disorders, cancers and others. The SNEYN may also be used for somatic
CC	or germline gene therapy.
XX	
SQ	Sequence 1992 BP; 671 A; 308 C; 372 G; 641 T; 0 other;
Alignment Scores:	
Pred. No.:	2,16e-232 Length: 1992
Score:	2437.00 Matches: 465
Percent Similarity:	100.00% Conservative: 0
Best Local Similarity:	100.00% Mismatches: 0
Query Match:	100.00% Indels: 0
DB:	22 Gaps: 0
US-09-744-313A-1 (1-465) x AAC86397 (1-1992)	
Qy	1 MetTyrLeuIleHisPheCysLeuIlePheArgAsnThrGlnLysArgGlyGluSerPhe 20
Dd	116 ATGTACTTGATCATTTTGTATTATATTCAGGAACACACAGAAAGGGGAGATCAITTT 175
Qy	21 GlyIleSerArgIleGlySerIysIleValPheLysSerThrThrMetGluGly 40
Dd	176 GGAATCAGCAGAATAGCTAGCAAAAATTAAGGCGATTCAAAGTAGTACCACAATGGAGGGA 235
Qy	41 AlaMetIeuProAsnTyrGlyValAlaGluGlyGluAspPheIleGluGlyIle 60
Dd	236 GCTATGTGCCTAATTATGCTGTACTGAAAGTGAAGATGATTTTATTTGAAGAAGGTATT 295
Qy	61 ValValMetGluAspSerProValGluAlaValSerThrProAsnThrProArgAsn 80
Dd	296 GTTGTAAATGGAAGATGATTTCTCATGTGAGGCTGTGAGCACACCTAATATCTCCCAGAAC 355
Qy	81 LeuAlaAlaTrpIyleSerIleProTyrValAspPheGluAspProSerSerGlu 100
Dd	356 CTTCGTGCATGGAATAATTAGCATTCATATGTAGACATTTTTTTTGGAGTCCCTCTCTGAA 415
Qy	101 ArgIysGluIysGluArgIleProValPheCysIleAspValGluArgAsnAspArg 120
Dd	416 AGGAAGAGAAAAAAGAAAGAAATTCCTGTGTTTGTATTGTATTGTGAAGAAATGATAGA 475
Qy	121 ArgAlaValGlyHisGluProGluHisIleTrpSerValTyrArgArgTyrLeuGluPheTyr 140
Dd	476 AGACATTTGGACACAGAGCTGAACATTCGTCTGTCTATAGAAATATCTTGAATTCAT 535
Qy	141 ValLeuGluSerIysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSer 160

DT	22-OCT-2001	(first entry)	
XX	XX	Human polynucleotide SEQ ID NO 1232.	
DE	XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;	
XX	XX	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	XX	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	XX	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	XX	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KW	XX	leukemia; ss.	
XX	XX	Homo sapiens.	
OS	XX	WO200153312-A1.	
PN	XX	26-JUL-2001.	
XX	XX	26-DEC-2000; 2000WO-US34263.	
XX	XX	21-JAN-2000; 2000US-0488725.	
PR	XX	25-APR-2000; 2000US-0552317.	
PR	XX	09-JUL-2000; 2000US-0598042.	
PR	XX	19-JUL-2000; 2000US-0620312.	
PR	XX	03-AUG-2000; 2000US-0653450.	
PR	XX	14-SEP-2000; 2000US-0662191.	
PR	XX	19-OCT-2000; 2000US-0693036.	
PR	XX	29-NOV-2000; 2000US-0727344.	
XX	XX	(HYSB-) HYSEQ INC.	
XX	XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	XX	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	XX	Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
XX	XX	WPI; 2001-442253/47.	
DR	XX	P-PSDB; AAM39873.	
XX	XX	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	XX	such as central nervous system injuries -	
XX	XX	Claim 1; SEQ ID NO 1232; 10078pp; English.	
XX	XX	The invention relates to human nucleic acids (AA157798-AA161369) and	
CC	XX	the encoded polypeptides (AM38642-AM42213) with nootropic, and	
CC	XX	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	XX	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	XX	of the invention may be used to treat diseases of the peripheral nervous	
CC	XX	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	XX	localised neuropathies and central nervous system diseases, such as	
CC	XX	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	XX	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	XX	utilisation of the activities such as: Immune system suppression,	
CC	XX	Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	XX	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	XX	assays for receptor activity, arthritis and inflammation, leukaemias and	
CC	XX	C.N.S disorders.	
CC	XX	Note: The sequence data for this patent did not form part of the printed	
CC	XX	specification.	
XX	XX	Sequence 1716 BP; 572 A; 297 C; 347 G; 500 T; 0 other;	
SQ	XX		
Alignment Scores:			
Pred. No.:	2,08e-227	Length:	1716
Score:	2386.00	Matches:	456
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	97.91%	Indels:	0
DB:	22	Gaps:	0
US-09-744-313A-1 (1-465) x AA159029 (1-1716)			
Qy	10	PheArgAsnThrGlnLysArgGlyGluSerPheGlyLysSerArgIleGlySerLysIle	29

Db	173	TTTCGGGACACACAGAAAGGGGAGAAATCATTTGGAAATCAGCAGATAGTAGCAAAAT	232
Qy	30	LysGlyValPheLysSerThrThrMetGluGlyValaMetLeuProAsnTyrGlyValAla	49
Db	233	AAAGGAGTATTCAAAGTACCAATGGAGGAGCTATGTTGCTTAATATGTTAGCT	292
Qy	50	GluGlyGluAspPheLleGluGlyIleValValMetGluAspSerProVal	69
Db	293	GAAGTGCAAGATGATTTTATGAAGAGGTATTGTTGTAATGGAAGATGATTCAGTG	352
Qy	70	GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro	89
Db	353	GAGCTGTGAGCACACCTAATCTCCCGAAACCTTGCTGCTGCAAAATTAGCATTC	412
Qy	90	TyrValAspPhePheGluAspProSerSerGluArgLysGluLysLysGluArgIlePro	109
Db	413	TATGTAGACTTTTGTGAGGATCCCTCTCTGAAGGAGGAGAGAGAGAGAGAGATTC	472
Qy	110	ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis	129
Db	473	GTGTTTTGTATTGATTTGAAGAAATGATAGAAGCAGTTCGACACAGCCTGAACAT	532
Qy	130	TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe	149
Db	533	TGGTCTGCTATAGAAGATATCTTGAATCTATGTACTTGAATCAAACTAAACAGAA	592
Qy	150	HisGlyValaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr	169
Db	593	CATGTCATTTCTGATGCCAGCTTCCTCTAAGAGGATCATTTGGCCCCCAAAATAT	652
Qy	170	GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuHis	189
Db	653	GAATTTCTTAAGTCAAGAGGAGAGAGTTCCTCAAGAGATATCTACAGAACTTCTG	712
Qy	190	ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr	209
Db	713	CCAGAACTGAGTAATAGTCAACTCTGGCAGACTTTCTTTCCCTTAATGTTGGGGAA	772
Qy	210	GlnPheLeuAspLysIleLeuProAsnValAsnLeuGlyLysIleIleLysSerValPro	229
Db	773	CAATTTCTTGAATAGTACTACAGAGATGTAATCTTTGGGAAATATATAAATCTGT	832
Qy	230	GlyLysLeuMetLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn	249
Db	833	GGAAACTAATGAAGAGAGAGTTCAGCTTGGACCTTTTATCATGATTTCAATAT	892
Qy	250	SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr	269
Db	893	TCITGTGAGTCTCCAAAGCCTAAACCAAGTAGACAGACTGACCATTTCTCAGCC	952
Qy	270	SerGluAsnLysLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGlu	289
Db	953	TCAGAAACCAACAGAGAGCTTTTCAATGATCTGTTTAAAAAATATGCAACCCG	1012
Qy	290	AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal	309
Db	1013	AATACAGAGAGAAAGCAAAATCAGAAATATTTTATGAGGTGATGCTGTGAGGAG	1072
Qy	310	TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHis	329
Db	1073	TATGATTAACCTGATGATGATGAGAGGGTATTTTCCAGGTTCTGCTGCTTCAT	1132
Qy	330	LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr	349
Db	1133	CTCTTAATGGAACTCGAACTCTCTTTAAACACCCCTGGAAATGTATCTATCTAT	1192
Qy	350	LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu	369
Db	1193	CTTCAGTGTAACCTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTCCTCA	1252
Qy	370	LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys	389
Db	1253	CTCAGAGATGCTATTTCTGTGAAACACCTGAACTCGCTCTCTCTCCAGATAGCA	1312



QY 270 SerGluAsnAsnLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
DB 1242 TCAGAAACACACAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTGTGAA 1183  
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
DB 1182 AATACAGAGAGAGAGACAAATCAGAAATATTTTATGAGGATGATGATGAGAGAGTC 1123  
QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrPheHis 329  
DB 1122 TATGATTACCTGATGATGATGAGAGGAGTATTTCCAGGTTCCCTGAGCTGCTCATCAT 1063  
QY 330 LeuLeuMetGlyThrArgLysLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
DB 1062 CTCCTTAATGGGAACCTCGAATCCTCTTTAAAAACACCTCGAATGATATATGATTAAT 1003  
QY 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369  
DB 1002 CTTCACTGTAACTAGAACAGCTATTTTCCAGGAGCCCGTTTGGTCTCACTCATACACTT 943  
QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
DB 942 CTCAGAGATGCTATATCTTGTGAAAACACTGAACCTCGCTCTCTCAAGATAGCAAAAA 883  
QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuValLysCys 409  
DB 882 GGAGCAAAACAGACTTTTGAAGAATGATGAATACATTCAGATCTGTAGTCAAGTGT 823  
QY 410 IleGlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnPro 429  
DB 822 ATTGGTGAGAACCAAGATATGAAGCATCAGACTCTCTGTTGATGGCTTACAGCACCA 763  
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
DB 762 GTACTACACAGAGAGCTGATTAATGTTTATTGGACATTTGTATACAGGAACCTGTTCCA 703  
QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
DB 702 GAGCTCAATAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 655

RESULT 4  
AAI60816/c  
ID AAI60816 standard; cDNA; 2176 BP.

AC AAI60816;  
XX  
XX 22-OCT-2001 (first entry)  
XX Human polynucleotide SEQ ID NO 4805.  
XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
XX leukaemia; ss.  
XX Homo sapiens.  
XX  
XX WO200153312-A1.  
XX  
XX 26-JUL-2001.  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
XX  
XX 21-JAN-2000; 2000US-0488725.  
XX 25-APR-2000; 2000US-0552317.  
XX 09-JUL-2000; 2000US-0598042.  
XX 19-JUL-2000; 2000US-0620312.  
XX 03-AUG-2000; 2000US-0653450.  
XX 14-SEP-2000; 2000US-0662191.  
XX 19-OCT-2000; 2000US-0693036.  
XX 29-NOV-2000; 2000US-0727344.  
XX

(HYSE-) HYSEQ INC.  
Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
WPI: 2001-442253/47.  
P-PSDB; AAM41660.  
Novel nucleic acids and polypeptides, useful for treating disorders  
such as central nervous system injuries -  
Claim 1; SEQ ID NO 4805; 10078pp; English.  
The invention relates to human nucleic acids (AAI57798-AAI61369) and  
the encoded polypeptides (AAM38642-AA42213) with nontropic,  
immunosuppressant and cytostatic activity. The polynucleotides are useful  
in gene therapy. A composition containing a polypeptide or polynucleotide  
of the invention may be used to treat diseases of the peripheral nervous  
system, such as peripheral nervous injuries, peripheral neuropathy and  
localised neuropathies and central nervous system diseases, such as  
Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
utilisation of the activities such as: Immune system suppression,  
Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
assays for receptor activity, arthritis and inflammation, leukaemias and  
C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
specification.  
SQ Sequence 2176 BP; 684 A; 427 C; 353 G; 712 T; 0 other;  
Alignment Scores:  
Pred. No.: 2,9e-227 Length: 2176  
Score: 2386.00 Matches: 456  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 97.91% Indels: 0  
DB: 22 Gaps: 0  
US-09-744-313A-1 (1-465) x AAI60816 (1-2176)  
QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29  
DB 2022 TTTCGGAACACACAGAAAAGGGAGAAATCATTTGGAATCAGCAGATAGTAGCAAAAT 1963  
QY 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
DB 1962 AAAGGAGTATTCAAAAGTACCACAAATGGAGGAGCTATGTTGCTTAATTATGGTGTAGCT 1903  
QY 50 GluGlyGluAspAspPheIleGluGluGlyValValMetGluAspAspSerProVal 69  
DB 1902 GAAGTGGAAGATGATTTTATTGAAGAGAGTATTGTTGTTATGAGAGATGATTCTCCAGTG 1843  
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaThrLysIleSerIlePro 89  
DB 1842 GAGGCTGTGACACACACTATACCTCCCGAAACCTTGTCTGATGAAAATTTAGCATTTCCA 1783  
QY 90 TyrValAspPhePheGluAspProSerSerGluArgLysGluLysLysGluArgIlePro 109  
DB 1782 TATGTAGACTTTTGTGAGGATCCCTCTCTGAAAGGAGGAGAAAAGAAAGAAATTCCT 1723  
QY 110 ValPheCysIleAspValGluArgAsnArgAlaValGlyHisGluProGluHis 129  
DB 1722 GTGTTTGTATGATGTTGAAAGAAATCATAGAGAGAGAGTTGGACAGCCCTGACAT 1663  
QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
DB 1662 TGGTCTGTCTATAGAGATATCTTGAATCTATGATCTTGAATCAAACTAACAGAAATTT 1603  
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169

Db 1602 CATGGTGCATTTCTTGATGCCAGCTTCTTCTAGAGGATCATTTGGCCCAAAATTTAT 1543  
Qy 170 GluPheLeuLysSerIysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHis 189  
Db 1542 GAATTTCTTAAGTCAAGAGGAGAGGTTCCAAGATATCTACAGAAATCTTCGAGCAT 1483  
Qy 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
Db 1482 CCAGAACTGAGTAATAGTCAACTTCTGGCAGACTTTCTTTCCCTTAATGTGGGAAACA 1423  
Qy 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLeuLysSerValPro 229  
Db 1422 CAATTTCTTGATAGATATCTACAGATGTAAATCTTTGGGAAATTTATAAATCTGTCT 1363  
Qy 230 GlyLysLeuMetLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
Db 1362 GGAACACTAATGAAGAGAGGTCAGCAATTTGGAACCTTTTATCATGAATTTCAATTA 1303  
Qy 250 SerCysGluSerProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
Db 1302 TCTTGTGAGTCTCCAAAGCCCTAAACCAAGTAGACCAACTGACCAATCTCAGCCCTACT 1243  
Qy 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGlu 289  
Db 1242 TCAGAAACCAACAGAGAGCTTTTCAATGATCTGTTTAAATATATCAACCGTGTGAA 1183  
Qy 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
Db 1182 AATACAGAGAGAGCAAAATCAGAAATTTATTTGAGGTGATGACTGTAGAGGAGTC 1123  
Qy 310 TyrAspTyrLeuMetTyrValGlyArgValPheGlnValProAspThrPheUHis 329  
Db 1122 TATGATTACCTGATGATAGGACGGGTAGTTTCCAGGTCTCTGACTGGCTTCATCAT 1063  
Qy 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
Db 1062 CTCTTAATGGGAATCGAACTCTTTTAAACACACCTCGAATATGATATGATTTACTAT 1003  
Qy 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369  
Db 1002 CTTGAGTGTAACTAGACAGCTATTTCAGGAGCACCGTTTGGTCTCCTCATACACTT 943  
Qy 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
Db 942 CTCAGAGATGCTATATTCTGTGAAACACACTGAACCTCGCTCTCTCCAGATAAGCAAAA 883  
Qy 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
Db 882 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTCAGATCTGTGTAGTCAAGTGT 823  
Qy 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
Db 822 ATTGTGTGAAGAAACCAAGATATGAAGCATCAGACTTCTGTTGTATGGCTTACAGCAACA 763  
Qy 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
Db 762 GTACTCAACAGCAGCTGACTATGTTTATTGGACATTTGTGATACAGGAATCTGTTCCA 703  
Qy 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
Db 702 GAGCTCAATAAGTACAAAGAGAGGTTACCTCTGTGACATCTGGATG 655

## RESULT 5

ABQ79518

ID ABQ79518 standard; cDNA; 3145 BP.

XX

AC ABQ79518;

XX

DT 25-NOV-2002 (first entry)

XX

DE cDNA encoding a protein similar to human sorting nexin.

XX

KW Atherosclerosis; antiarteriosclerotic; marker; cardiovascular;  
KW SSH 6; gene; human; nexin; ss.

XX Homo sapiens.

OS

XX Key Location/Qualifiers

FH 128..2968

FT /\*tag= a

XX

PN W0200262839-A2.

XX

PD 15-AUG-2002.

XX 05-FEB-2002; 2002WO-EP01327.

XX 07-FEB-2001; 2001EP-0200439.

XX (UTMA-) UNIV MAASTRICHT.

XX Daamen MJAP, Cleutjens CBJM, Zaman GJR;

XX WPI; 2002-643400/69.

XX P-PSDB; ABB81193.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Use of a polynucleotide differentially expressed in ruptured and stable atherosclerotic plaques as a marker for atherosclerosis, useful in treating, diagnosing or preventing atherosclerosis -

Example 6; Page 35-37; 4pp; English.

The invention relates to the use of a polynucleotide differentially expressed in ruptured and stable atherosclerotic plaques as a marker for atherosclerosis, where the polynucleotides can be selected from the sequences shown in ABQ79517-19. The polynucleotides are useful as a marker of atherosclerosis, which may be used: (i) in the diagnosis, prevention and treatment of atherosclerosis; (ii) as serum/plasma markers to screen patients at risk for plaque instability to evaluate the effects of other treatments; (iii) in the preparation of vector molecules for the expression of the encoded protein in host cells; and (iv) in the identification of functional targets or analogues of the gene. The polynucleotides, the encoded proteins or antibodies against the proteins may be used to target other therapeutics to an unstable plaque. Modulation of the expression of the polynucleotide can increase plaque stability and therefore inhibit the progression of atherosclerotic cardiovascular disease. Modulators may be used to prepare pharmaceuticals for atherosclerotic disorders. The present sequence represents a cDNA encoding a protein similar to human sorting nexin.

Sequence 3145 BP; 985 A; 570 C; 659 G; 931 T; 0 other;

## Alignment Scores:

Pred. No.: 1.21e-226 Length: 3145  
Score: 2382.00 Matches: 454  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 99.56% Mismatches: 0  
Query Match: 97.74% Indels: 0  
DB: 24 Gaps: 0

US-09-744-313A-1 (1-465) x ABQ79518 (1-3145)

Qy 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerIysIle 29  
Db 1598 TTTCGGAACACACAGAAAGGGAGAAATCAATTTGGAAATCAGCAGATAGTAGCAAAATT 1657

Qy 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
Db 1658 AAAGGAGTATTCAGAGTACCACATATGGAGGGAGCTATGTTGCCATTAATGCTAGCT 1717

Qy 50 GluGlyGluAspAspPheIleGluGlyIleValMetGluAspAspSerProVal 69  
Db 1718 GAAGGTGAAGATGATTTTATTGAAAGAGGTATTGTTGTAATGAAGATGATCTCCAGTG 1777

Qy 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro 89



	Ddb	1778	GAGGCTGTGACACACCTTAATACTCCCGAAACCCTTGTGCATGGAAAATTAGCATTTCCA	1837
	Oy	90	TyrValAspPhePheGluAspProSerSerGluArgLysGluLysLysGluArgIlePro	109
	Ddb	1838	TATGTAGA CTTTTTTTGAGGATCCCTTCCTGTGAAGAGGAGGAAAAAAGAAGAAATTCCT	1897
	Oy	110	ValPheCysIleAspValGluArgAsnAspArgAlaValGLYHisIleGluProGluHis	129
	Ddb	1898	GTGTTTTGTATTGATGTGTAAGAANAATGATAGAGAGCAAGTTGGACACGAGCCTGNAACAT	1957
	Oy	130	TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe	149
	Ddb	1958	TGGTCTGCTATAAGATATCTTGAAATTCATATGTACTTGAATCAAACTAACAGAAATTT	2017
	Oy	150	HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysSerTyr	169
	Ddb	2018	CATGCTGCATTTCCGTGATGCCAGCTTCCTCTTAAGAGATCATTTGGCCCCAAAAATTAT	2077
	Oy	170	GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGluLysLeuLeuGlnHis	189
	Ddb	2078	GAATTTCTTAAGTCAAAGAGGAGAGATTCAGAGATATCTACAGAACTTCTGCAGCAT	2137
	Oy	190	ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr	209
	Ddb	2138	CCAGAACTGAGTAATAGTCAACTTCTGGCAGACTTCTTTCCCTAAATGCTGGGAAAAACA	2197
	Oy	210	GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro	229
	Ddb	2198	CAATTTCTTGATAAGATATACTACAGATGTAAATCTTTGGGAAAATATATAAAATCTGTCTCT	2257
	Oy	230	GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPhelLeuAsn	249
	Ddb	2258	GGAAACTAATGAAAGAGAAAGGTGAGCAATTTGGAACCTTTTATCATGATTTTCATTANT	2317
	Oy	250	SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr	269
	Ddb	2318	TCTTTGTGAGTCTCCAAAGCCTAAACCAGTAGACCAGAACTGACCACTTCTCACGCCCTACT	2377
	Oy	270	SerGluAsnAsnLysLysIlePheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu	289
	Ddb	2378	TCAGAAAAACAACAAGAGCTTTTCAATGATCTGTGTTAAAAATAAATCAACACCGTGTCTGAA	2437
	Oy	290	AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal	309
	Ddb	2438	AATACAGAGAGAAAGCAAAATCAGAAATTTATTTATGGAGGTGATGACTGTAGNAGAGATC	2497
	Oy	310	TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHis	329
	Ddb	2498	TATGATTAACCTGATGTATGTAGACGGGTAGTTTCCAGATTCTCGACTGGCTTCATCAT	2557
	Oy	330	LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr	349
	Ddb	2558	CTCTTAATGGGAACTCGAATCCTCTTTAAAAAACACCCCTGGAAATGTATATCTGATTACTAT	2617
	Oy	350	LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu	369
	Ddb	2618	CTTCAGTGTAAACTAGAACAGCTATTTACAGAGCACCGTTTGGTCTCACTCATPAACACTT	2677
	Oy	370	LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys	389
	Ddb	2678	CTCAGAGATGCTATATTTCTGTGAAAAACACTGAACTCGCTCTCTCCAGATAAGCAAAAA	2737
	Oy	390	GlyAlaLysGlnThrPheGluGluMetSerAsnTyrIleProAspLeuLeuValLysCys	409
	Ddb	2738	GGAGCAAAAACAGACTTTTGAAGAANAATGATGAATTTACATTCAGATCTGTGTAGTCAAGTGT	2797
	Oy	410	IleGlyGlnGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro	429
	Ddb	2798	ATTGGTGAGGAAAAACAAGTATGAAGACATCAGATCTTCTGTTGATGGCTTACAGCAACCA	2857
	Oy	430	ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro	449

Db	2858	GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTTGTATACAGGAACTGTTTCCA	2911
Qy	450	GIuLeuAaLnLySValGlnLySGLuValThrSerValThrSerTrpMet	465
Db	2918	GAGTCATATAGGTACAAAGGAAGTTACTCTCTGTGACATCTTGGATG	2965
Db	2918	GAGTCATATAGGTACAAAGGAAGTTACTCTCTGTGACATCTTGGATG	2965
RESULT 6			
AAI59030			
ID	AAI59030	standard; cDNA; 1551 BP.	
XX	AC	AAI59030;	
XX	DT	22-OCT-2001 (first entry)	
XX	XX	Human polynucleotide SEQ ID NO 1233.	
XX	XX	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;	
KW	KW	peripheral nervous system; neuropathy; central nervous system; CNS;	
KW	KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;	
KW	KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;	
KW	KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;	
KW	KW	leukemia; ss.	
XX	XX	Homo sapiens.	
OS	OS	WO200153312-A1.	
PN	PN	26-JUL-2001.	
XX	XX	26-DEC-2000; 2000WO-U834263.	
XX	XX	21-JAN-2000; 2000US-0488725.	
PR	PR	25-APR-2000; 2000US-0552317.	
PR	PR	09-JUL-2000; 2000US-0598042.	
PR	PR	19-JUL-2000; 2000US-0620312.	
PR	PR	03-AUG-2000; 2000US-0653450.	
PR	PR	14-SEP-2000; 2000US-0662191.	
PR	PR	19-OCT-2000; 2000US-0693036.	
PR	PR	29-NOV-2000; 2000US-0727344.	
XX	XX	(HYSE-) HYSEQ INC.	
XX	XX	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;	
PI	PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;	
PI	PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;	
XX	XX	WPI; 2001-442253/47.	
DR	DR	P-PSDB; AAM39674.	
XX	XX	Novel nucleic acids and polypeptides, useful for treating disorders	
PT	PT	such as central nervous system injuries -	
XX	XX	Claim 1; SEQ ID NO 1233; 10078pp; English.	
PS	PS	The invention relates to human nucleic acids (AAI57798-AAI61369) and	
XX	XX	the encoded polypeptides (AAM38642-AAM42213) with nootropic,	
CC	CC	immunosuppressant and cytostatic activity. The polynucleotides are useful	
CC	CC	in gene therapy. A composition containing a polypeptide or polynucleotide	
CC	CC	of the invention may be used to treat diseases of the peripheral nervous	
CC	CC	system, such as peripheral nervous injuries, peripheral neuropathy and	
CC	CC	localised neuropathies and central nervous system diseases, such as	
CC	CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic	
CC	CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the	
CC	CC	utilisation of the activities such as: Immune system suppression,	
CC	CC	activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic	
CC	CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,	
CC	CC	assays for receptor activity, arthritis and inflammation, leukaemias and	
CC	CC	C.N.S disorders.	
CC	CC	Note: The sequence data for this patent did not form part of the printed	
CC	CC	specification.	
XX	XX	Sequence 1551 BP; 527 A; 258 C; 317 G; 449 T; 0 other;	

Alignment Scores:			
Pred. No.:	8,848-195	Length:	1551
Score:	2057.50	Matches:	401
Percent Similarity:	87.94%	Conservative:	0
Best Local Similarity:	87.94%	Mismatches:	0
Query Match:	84.43%	Indels:	55
DB:	22	Gaps:	1
US-09-744-313A-1 (1-465) x AAI59030 (1-1551)			
QY	10	PheArgAsnThrGlnLysArgGlyGluSerPheGlyLeuSerArgIleGlySerLysIle	29
Db	173	TTTCGGAACACACAGAAAAGGGAGAAATCATTTGGAAATCAGAGAAATAGTAGCAAAAT	232
QY	30	LysGlyValPheLysSerThrMetGluGlyValMetLeuProAsnTyrGlyValAla	49
Db	233	AAAGAGATATTCAAAGTACCAATAGGAGGAGCTATGTTGCCTAAATATGCTGTAGCT	292
QY	50	GluGlyGluAspPheIleGluGlyValValMetGluAspSerProVal	59
Db	293	GAAGGTGAAGATGATTTTATGAAGAAGGTATTTGTTGTAATGGAAGATGATTTCCAGTG	352
QY	70	GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro	89
Db	353	GAGGCTGTGAGCACACCTAATATCTCCCGAAACCTTGCTGCATGGAAATTTAGCATTC	412
QY	90	TyrValAspPheGluAspProSerSerGluArgLysGluLysGluArgIlePro	109
Db	413	TATGTAGATTTTTTTGAGATTCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGG	472
QY	110	ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis	129
Db	473	GTGTTTTGTATTGATTTGAAGAAATGATAGAGAGCAGTTGGACACAGCCTGAACAT	532
QY	130	TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe	149
Db	533	TGGTCTGTCTATAGATATCTTTGAATTTCTATGTACTTGAATCAAACTAAACAGATTT	592
QY	150	HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleLeuGlyProLysAsnTyr	169
Db	593	CATGGTGCATTTCTGTAGTCCAGCTTCTTCTTAAGAGATCATTTGCCCCCAAAATAT	652
QY	170	GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis	189
Db	653	GAATTCCTTAAGTCAAGAGGAGAGGATTTCCAGATATCTACAGAACTTCTGCAGCAT	712
QY	190	ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr	209
Db	713	CCAGAACTGAGTAAATAGTCAACTCTCTGCGACACTTTCTTCCCTAATATGTTGGGAA	772
QY	210	GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLeuLysSerValPro	229
Db	773	CAATTTCTTTGATAGATATCTACAGATGTAATCTTTGGGAAATATATAAATCTGTCT	832
QY	230	GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn	249
Db	833	GGAAACTAATGAAGAGAGAGGTCAGCATTTTGGACCTTTTATCATGAAATTCATTAAT	892
QY	250	SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr	269
Db	893	TCTTTGTGAGTCTCCAAAGCCTAAACCAAGTAGACAGCAACTGACCATTTCTCAGCCCT	952
QY	270	SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu	289
Db	953	TCAGAAAAACACAGAAAGCTTTTCAATGATCTCTTTTAAAAATTAATGCAACCGCTGAA	1012
QY	290	AsnThrGluArgGlyGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal	309
Db	1013	AAATACAGAGAGAAAGCAAAATCAGAAATATTTTATGGAGGTGATGACCTGAGAGGATC	1072
QY	310	TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHisHis	329
Db	1073	TATGATTAATCTGATGATGTA-----	1093
QY	330	LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr	349
Db	1093	-----	1093
QY	350	LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu	369
Db	1093	-----	1093
QY	370	LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys	389
Db	1094	-----GATGCTATATCTGTGAAACACATGAACCTCGCTCTCTCCAGATAGCAAAAA	1147
QY	390	GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys	409
Db	1148	GGACCAAAACAGACTTTTGAAGAAATGATGAAATACATTCAGATCTGTGTAGTCAAGTGT	1207
QY	410	IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro	429
Db	1208	ATTGCTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTGTTGATGGCTTTACAGCAACA	1267
QY	430	ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro	449
Db	1268	GTACTCAACAGCAGCTGACTTATGTTTATTTGGACATTGTGATACAGGAATGTTTCCA	1327
QY	450	GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet	465
Db	1328	GAGTCAATAGGTACAAAAGGAAGTTACTCTGTGACATCTTGGATG	1375
RESULT 7			
ID	AAZ17379	standard; cDNA; 779 BP.	
XX	AAZ17379;		
XX	12-OCT-1999 (first entry)		
DE	Human gene expression product cDNA sequence SEQ ID NO:4852.		
XX	Human; gene; gene expression product; diagnosis; therapy; probe;		
KW	detection; mapping; tissue typing; profiling; forensic; cancer;		
KW	genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.		
XX	Homo sapiens.		
XX	WO9938972-A2.		
XX	05-AUG-1999.		
XX	28-JAN-1999; 99WO-US01619.		
XX	03-APR-1998; 98US-0080666.		
PR	28-JAN-1998; 98US-0072910.		
PR	24-FEB-1998; 98US-0075954.		
PR	31-MAR-1998; 98US-0080114.		
PR	03-APR-1998; 98US-0080515.		
XX	(CHIR ) CHIRON CORP.		
PA	(HYSE-) HYSEQ INC.		
XX	Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;		
PI	Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;		
PI	Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;		
PI	Lamson G, Leskowitz D, Pot D, Randazzo F, Reinhard C;		
PI	Stache-Crain B, Sudduth-Klinger J, Williams LT;		
XX	WPI; 1999-494092/41.		
XX	Novel human genes and their expression products which are		
PT	differentially expressed in different cell types		
XX	Claim 1; Page 2307; 2479pp; English.		
PS			
XX			

The present invention describes a library of human polynucleotides comprising the sequences given in AA121532 to AA217779. Also described is a method of detecting differentially expressed genes correlated with the cancerous state of a mammalian cell, comprising detecting at least one differentially expressed gene product in a test sample from a cell suspected of being cancerous, where the gene product is encoded by one of the 5248 polynucleotide sequences given in AA121532 to AA217779. The polynucleotides can be used as a source of primers and probes, which can be used for a variety of purpose, e.g. detection of expression levels, mapping, tissue typing or profiling, forensics, genetic analysis and detection of polymorphisms. Polypeptides encoded by the polynucleotides can be used for raising antibodies for experimental, diagnostic and therapeutic purposes. The polynucleotides may also be used to construct arrays for diagnostics (which may be used to determine function of an encoded protein); and to detect differences in expression levels between two cells (e.g. to identify abnormal or diseased tissue in a human, to identify a genetic predisposition or susceptibility to a disease such as cancer). The polynucleotides of the invention are especially used in the diagnosis, prognosis and management of colorectal cancer, breast cancer, and lung cancer. The polynucleotides can also be used to screen for peptide analogues and antagonists.

Sequence 779 BP; 248 A; 143 C; 153 G; 232 T; 3 other; XX \$0

Alignment Scores:		
Pred. No.:	2.44e-95	779
Score:	1054.00	205
Percent Similarity:	99.04%	Conservative: 2
Best Local Similarity:	99.09%	Mismatches: 1
Query Match:	43.25%	Indels: 1
DB:	20	Gaps: 0

US-09-744-313A-1 (1-465) x AAZ17379 (1-779)

257	QY	LysProSerArgProGlu-LeuThrIleLeuSerProThrSerGluAsnAsnLysLe	276
...	...	...	...
61	Db	AGACCAAGTAGACCAGAAACTGACCATTCCTCAGTCTACTTCAGAAACAACAAGAGCT	120
276	QY	upheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAs	296
121	Db	TTTCAATGATCTGTTTAAAAATAATCAAAACCGTCTGAAATAATCAGAGAGAAAGCAAAA	180
296	QY	nglnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrVa	316
181	Db	TCAGAAATATTATTTGAGAGGTGATGACTGTAGAAGGAGTCTATGATTACCTGGATGTATGT	240
316	QY	lglYArgValValPheGlnValProAspTyrLeuHisIleuLeuMetGlyThrArgIl	336
241	Db	AGGACGGGTAGTGTTCACAGGTCCCTGACTGGCTTCATCATCTCTTAATGGGAACTCGAAT	300
336	QY	eLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGl	356
301	Db	CCTCTTTAAAAACACCCCTGGAAATGTATCTGATTACTATCTTCAGTGTAAACTAGAAC	360
356	QY	nLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCy	376
361	Db	GCTATTTTCAGGAGCACCGTTGGTCTCACTCAATAACATCTTCACAGATGCTATATTCTG	420
376	QY	sGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGl	396
421	Db	TGAAACACACTGAACCTCGCTCTCTCAAGATTAAGCAAAAGGAGCAAAACAGACTTTTGA	480
396	QY	uGluMetMetAsnTyrIleProAspLeuValLysCysIleGlyGluGluThrLysTy	416
481	Db	AGAAATGATGTAATACATTCAGATCTGTATGTCAGTGTATTTGTGTAAGAAACCAAGTA	540
416	QY	rgluSerIleArgLeuLeuPheAspGlyLeuGlnProValLeuAsnLysGlnLeuTh	436
541	Db	TGAAAGCATTCAGACTTCTGTTTGTATGGCTTACAGCAACAGTACTCTCAACAGCAGCTGAC	600
436	QY	rtYrValLeuLeuAspIleValIleGlnLeuPheProGluLeuAsnLysValGlnLys	456
601	Db	TTATGTTTATTGGACATTTGTGATCAGGAACGTGTTTNCAGAGCTCAATAAGGTCAAAA	660

Qy	456	sGluValThrSerValThrSerTTP	464
Dd	661	GGAAGTTACCTCTGTGCATCTTGG	685
 RESULT 8 ABS77486			
ID	ABS77486	standard; cDNA; 725 BP.	
XX	AC	ABS77486;	
XX	DT	12-DEC-2002 (first entry)	
XX	DE	Prog embryonic gene sequence Q9925894.	
XX	KW	Frog; ss; embryonic development; developo-	
KW	KW	microarray; cell differentiation.	

XX Xenopus laevis.  
OS  
XX  
XX  
PN US2002081610-A1.  
XX  
XX  
PD 27-JUN-2002.  
XX  
XX 23-JUL-2001; 2001US-0910943.  
XX  
XX 21-JUL-2000; 2000US-219658P.  
XX  
XX (UYRQ ) UNIV ROCKEFELLER.  
XX  
XX Hemmati-Brivanlou A, Altmann CR;  
PI  
XX  
XX WPI; 2002-626534/67.  
DR  
XX  
XX Nucleic acid array containing Xenopus embryonic nucleic acids is useful  
PT to identify genes involved in embryonic development. to identify  
PT different types of embryonic cells, and to diagnose developmental  
PT disorders -  
PT

XX  
PC  
Claim 1: page 756: English: 823pp:

XX The invention relates to a nucleic acid array, where each coordinate  
CC contains a single nucleic acid species having one of 770 nucleotide  
CC sequences (appearing as ABS76747-ABS77516) of a *Xenopus* embryonic gene  
CC product, or its complement or hybridisable fragment of not less than  
CC 20 contiguous nucleotides of one of those sequences. Also included are  
CC detecting differential expression of embryonic genes, comprising:  
CC (a) contacting a nucleic acid array comprising genes expressed in  
CC embryonic but not mature cells with nucleic acids from sample and  
CC control cells; and (b) detecting differential hybridisation of nucleic  
CC acids from the sample cells relative to the control cells; and detecting  
CC defects in development, comprising: (a) contacting nucleic acids from  
CC test cells undergoing development with a nucleic acid array of gene  
CC products known to play a fundamental role in the development process; and  
CC (b) detecting a difference in expression of a fundamental gene in the  
CC sample cells relative to a standard. The invention is useful to identify  
CC genes involved in embryonic development and related processes such as  
CC cell differentiation. This would be useful for diagnosing developmental  
CC disorders and for identifying different types of embryonic cells.  
CC The present sequence is one of the 770 *Xenopus* embryonic cDNA sequences.

AX  
SQ Sequence 725 BP; 229 A; 135 C; 140 G; 218 T; 3 other;

**Alignment Scores:**

Alignment Scores:		
Pred. No.:	3,62e-88	Length:
Score:	981.50	Matches:
Percent Similarity:	85.83%	Conservative:
Best Local Similarity:	80.42%	Mismatches:
Query Match:	40.27%	Indels:
DB:	24	Gaps:
		1
		725

US-09-744-313A-1 (1-465) X ABS77486 (1-725)

187 LeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGly 206  
20 TTGCAGGATCCCATCGAATTCGAATTCGTCGACCCACGCG-----TCGGGT 64  
207 GlyGluThrGlnPheLeuAspLysLeuProAspValAsnLeuGlyLysLeuLys 226  
65 TTGGAACACTCAGTTTCATGACAAATCCACAGATGTGAATCTGGGAAATGATCAAG 124  
227 SerValProGlyLysLeuMetLysGlyGlnHisLeuGluProPheLeuMetAsn 246  
125 TCGATTCCTGGAAACTCATTAAGAAAGGCGACATCTGGAGCTGTTTATATGAT 184  
247 PheLeuAsnSerCysGluSerProLysProLysProSerArgProGluLeuThrLeu 266  
185 TTCATTAACATCATGTGAATCACCACCAACCAACCAAGTAGGCTGAACTGACTATTCTA 244  
267 SerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsn 286  
245 AGCCCGACTTCAGAAACCAATAAAGCTTTTAAATGATCTGTACAGAAATATGCTAAT 304  
287 ArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnThrPheMetGluValMetThrVal 306  
305 CGTCTCGAAATACAGAAAGGAGCATACCAAGATATCTTTATGGAAGTGTACCGTA 364  
307 GluGlyValThrAspTyrLeuMetTyrValGlyArgValPheGlnValProAspTrp 326  
365 GAGGGTGTATGACTATTATATGATATAGGACGCTGTGTTTTCACATCTCGATTGG 424  
327 LeuHisLeuLeuMetGlyThrArgLysLeuPheLysAsnThrLeuGluMetTyrThr 346  
425 TTTTCATCATCTTTGATGGGTGACGAATCTCTTTAAACATACACCTCGAGACTTACACA 484  
347 AspTyrTyrLeuGlnCysLysLeuGluGlnPheGlnHisArgLeuValSerLeu 366  
485 AATAGCTATTAACTATAATTAAGACGCTCTTTCAAGAGCATCGCTGTTTCTTTG 544  
367 IleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAsp 386  
545 ATTACTCTGCTTAGAGATGCTATATTGTAACCGCTGACACGATCATACATCG 604  
387 LysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeu 406  
605 AAGCAGCAAGGGGCAAGCTTACATTTGAAGAAATGATGCGCTATATTCAGATTGGAT 664  
407 ValLysCysIleGlyGluGluThrLysTyrGluSerIleArgLeuPheAspGlyLeu 426  
665 GGTAATATGTTTGTGTGATGAAGCTAAATATGAAGGCATGCGACTTCTGTTTGTGACTG 724

RESULT 9  
ABX50611  
ID ABX50611 standard; cDNA; 402 BP.  
XX  
AC ABX50611;  
XX  
AC  
XX  
DT 25-FEB-2003 (first entry)  
XX  
DE Bovine EST associated with lactation/muscle/fat deposition #540.  
XX  
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
KW muscle deposition; fat deposition; genome mapping; gene identification;  
KW gene analysis; cattle breeding.  
XX  
OS Bos Taurus.  
XX  
FN US2002137160-A1.  
XX  
PD 26-SEP-2002.  
XX  
PF 26-OCT-2001; 2001US-0983965.  
XX  
PR 17-DEC-1998; 98US-113678P.  
PR 15-DEC-1999; 99US-0455231.  
XX

PA (BYAT/) BYATT J C.  
PA (MATH/) MATHIALAGAN N.  
PA (TAON/) TAO N.  
PA (WARR/) WARREN W C.  
XX  
XX Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX WPI; 2003-102386/09.  
XX  
XX Purified nucleic acid molecules, useful for genome mapping, gene  
XX identification and analysis, cattle breeding or preparation of  
XX constructs for cattle gene expression and genetically improved cattle -  
XX  
XX Claim 2; SEQ ID No 540; 38pp; English.  
XX  
XX The invention relates to a purified nucleic acid molecule associated with  
XX lactation or muscle and fat deposition (designated LMFD), derived  
XX from cattle, and the LMFD nucleic acid can specifically hybridise to a  
XX second nucleic acid molecule comprising any of 5912 nucleotide  
XX sequences, appearing as ABX50072-ABX5983, or complements of them.  
XX Also included are; (1) a transformed cell having a nucleic acid  
XX comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
XX translated sequence that functions in the cell to cause termination of  
XX transcription and addition of polyadenylated ribonucleotides to a 3' end  
XX of the mRNA molecule; and (2) determining a level or pattern of a  
XX molecule in a bovine cell or tissue comprising: (a) incubating a marker  
XX nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
XX complement or fragment) with a complementary nucleic acid molecule  
XX obtained from the bovine cell or tissue, where hybridisation between the  
XX marker nucleic acid and the complementary nucleic acid permits the  
XX detection of the molecule; and (b) detecting the level or pattern of the  
XX complementary nucleic acid, where the detection of the complementary  
XX nucleic acid is predictive of the level or pattern of the molecule.  
XX The LMFD nucleic acid is used for determining a level or pattern  
XX of a molecule in a bovine cell or tissue. It is useful for genome  
XX mapping, gene identification and analysis, cattle breeding, preparation  
XX of constructs for use in cattle gene expression, or for genetically  
XX improving cattle. The present sequence is one of the 5912 bovine  
XX LMFD EST (expressed sequence tag) nucleic acids.  
XX Note: The present sequence was not shown in the specification but  
XX was obtained in electronic format from the USPTO web site:  
XX seqdata.uspto.gov/sequence.html?DocID=20020137160.  
XX  
XX Sequence 402 BP; 141 A; 71 C; 76 G; 113 T; 1 other;  
XX  
XX Alignment Scores:  
XX Pred. No.: 1.01e-58 Length: 402  
XX Score: 682.00 Matches: 127  
XX Percent Similarity: 99.25% Conservative: 5  
XX Best Local Similarity: 95.49% Mismatches: 1  
XX Query Match: 27.99% Indels: 0  
XX DB: 25 Gaps: 0  
XX  
XX US-09-744-313A-1 (1-465) x ABX50611 (1-402)  
XX  
XX QY 205 AsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLys 224  
XX Db 3 AATGGTGGGAAACACAGCTTCTTGATAAGATACTACCAGACGTAATCTTGGGAAAT 62  
XX  
XX QY 225 IleLysSerValProGlyLysLeuMetLysGlyGlnHisLeuGluProPheLeu 244  
XX Db 63 ATAAATCTGTTCTCTGAAAACTAATGAAAGAGAAAGTCAACATTTTGGACCTTCATC 122  
XX  
XX QY 245 MetAsnPheIleAsnSerCysGluSerProLysProLysProSerArgProGluLeuThr 264  
XX Db 123 ATGAATTTTCATTAATTTCTTGAAATCTCTCAAGGCTTAACCGATTAACCGAACTGACC 182  
XX  
XX QY 265 IleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsn 284  
XX Db 183 ATTCTACCCCTTACTCTAGAGATTAATAGAGCTTTTATGATCTGTATAGATAAT 242  
XX  
XX QY 285 AlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnThrPheMetGluValMet 304  
XX

Db 243 GCRAACCGTCTGAGAAATACAGAAAGGCAAAATCAGAAATATTTATGGAATGATG 302  
Qy 305 ThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValPro 324  
Db 303 ACTGTAGAAGGAGTCTATGATTTACCTGATGTAAGTANGACAGTGGTTTCCAGTTCT 362  
Qy 325 AspTyrLeuHisLeuLeuMetGlyThrArgLeu 337  
Db 363 GACTGGCTTCATCATCTCTTAATGGAAGTCAATCTTC 401

RESULT 10  
AAC02755  
ID AAC02755 standard; cDNA; 358 BP.  
AC AAC02755;  
XX  
DT 06-OCT-2000 (first entry)  
XX  
DE Human secreted protein 5' EST, SEQ ID NO: 2753.  
XX  
KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
KW gene therapy; chromosome mapping; ss.  
XX  
OS Homo sapiens.  
XX  
PN EP1033401-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 21-FEB-2000; 2000BP-0200610.  
XX  
PR 26-FEB-1999; 99US-0122487.  
XX  
PA (GEST ) GENSET.  
XX  
PI Dumas Milne Edwards J, Duclert A, Giordano J;  
XX  
DR WPI; 2000-500381/45.  
DR P-PSDB; AAG02749.  
XX  
PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
XX  
XX Claim 1; SEQ ID 2753; 71pp + CD-ROM; English.  
XX  
CC The present sequence is one of a large number of 5' ESTs derived from  
CC mRNAs encoding secreted proteins. An ORF has been identified within the  
CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs  
CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC mRNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.  
XX  
SQ Sequence 358 BP; 125 A; 63 C; 69 G; 101 T; 0 other;

Alignment Scores:  
Pred. No.: 5.15e-54 Length: 358  
Score: 634.00 Matches: 119  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 26.02% Indels: 0  
DB: 21 Gaps: 0

US-09-744-313A-1 (1-465) x AAC02755 (1-358)

Qy 216 LeuProAspValAsnLeuGlyLysIleLeuLysSerValProGlyLysLeuMetLysGln 235  
Db 2 CTACCAGATCTAAATCTTGGAAAAATATAAAATCTGTTCTCTGAAAACTAATGAAAGAG 61  
Qy 236 LysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLys 255  
Db 62 AAAGGTGAGCATTTGGAAACCTTTTATCATGAATTTCAATTAATCTTTGTGAGTCTCAAAG 121  
Qy 256 ProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLys 275  
Db 122 CTTAAACCAAGTAGACAGCAACTGACCACTTCTCAGCCCTACTTCAGAAACACAGAGAG 181  
Qy 276 LeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGln 295  
Db 182 CTTTTCATGATCTGTTTAAAAATAATGCAACCGTGTGAAATATACAGAGAGAAAGCAA 241  
Qy 296 AsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyr 315  
Db 242 AATCAGAAATATTTATGAGAGTGTGACTGTAGAAGGAGTCTATGATTACCTGTAT 301  
Qy 316 ValGlyArgValValPheGlnValProAspTyrLeuHisLeuLeuMetGlyThr 334  
Db 302 GTAGGACGGGTAGTTTTCAGGTTCCTGACTGCTTCATCATCTCTTAATGGGAACC 358

RESULT 11  
ABX50426  
ID ABX50426 standard; cDNA; 305 BP.  
XX  
AC ABX50426;  
XX  
DT 25-FEB-2003 (first entry)  
XX  
DE Bovine EST associated with lactation/muscle/fat deposition #355.  
XX  
KW Bovine; ss; EST; expressed sequence tag; lactation; LMFD;  
KW muscle deposition; fat deposition; genome mapping; gene identification;  
KW gene analysis; cattle breeding.  
XX  
OS Bos Taurus.  
XX  
PN US2002137160-A1.  
XX  
PD 26-SEP-2002.  
XX  
PP 26-OCT-2001; 2001US-0983965.  
XX  
PR 17-DEC-1998; 98US-113678P.  
PR 15-DEC-1999; 99US-0465231.  
XX  
PA (BYAT//) BYATT J C.  
PA (MATH//) MATHIALAGAN N.  
PA (TAON//) TAO N.  
PA (WARR//) WARREN W C.  
XX  
PI Byatt JC, Mathialagan N, Tao N, Warren WC;  
XX  
XX WPI; 2003-102386/09.  
XX  
PT Purified nucleic acid molecules, useful for genome mapping, gene  
PT identification and analysis, cattle breeding or preparation of  
PT constructs for cattle gene expression and genetically improved cattle -  
XX  
PS Claim 2; SEQ ID No 355; 38pp; English.  
XX  
CC The invention relates to a purified nucleic acid molecule associated with  
CC lactation or muscle and fat deposition (designated LMFD), derived to a  
CC from cattle, and the LMFD nucleic acid can specifically hybridize to a  
CC second nucleic acid molecule comprising any of 5912 nucleotide  
CC sequences, appearing as ABX50072-ABX55983, or complements of them.  
CC Also included are: (1) a transformed cell having a nucleic acid  
CC comprising an LMFD nucleic acid linked to a promoter and a 3' non-  
CC translated sequence that functions in the cell to cause termination of  
CC transcription and addition of polyadenylated ribonucleotides to a 3' end

CC of the mRNA molecule; and (2) determining a level or pattern of a  
CC molecule in a bovine cell or tissue comprising: (a) incubating a marker  
CC nucleic acid (comprising any of the 5912 nucleic acid sequences or its  
CC complement or fragment) with a complementary nucleic acid molecule  
CC obtained from the bovine cell or tissue, where hybridization between the  
CC marker nucleic acid and the complementary nucleic acid permits the  
CC detection of the molecule; and (b) detecting the level or pattern of the  
CC complementary nucleic acid, where the detection of the complementary  
CC nucleic acid is predictive of the level or pattern of the molecule.  
CC The LMFD nucleic acid is used for determining a level or pattern  
CC of a molecule in a bovine cell or tissue. It is useful for genome  
CC mapping, gene identification and analysis, cattle breeding, preparation  
CC of constructs for use in cattle gene expression, or for genetically  
CC improving cattle. The present sequence is one of the 5912 bovine  
CC LMFD EST (expressed sequence tag) nucleic acids.  
CC Note: The present sequence was not shown in the specification but  
CC was obtained in electronic format from the USPTO web site:  
CC seqdata.uspto.gov/sequence.html?DocID=20020137160.  
XX

SQ Sequence 305 BP; 109 A; 60 C; 51 G; 85 T; 0 other;

Alignment Scores:  
Pred. No.: 2,83e-41 Length: 305  
Score: 505.00 Matches: 97  
Percent Similarity: 97.03% Conservations: 1  
Best Local Similarity: 96.04% Mismatches: 3  
Query Match: 20.72% Indels: 0  
DB: 25 Gaps: 0

US-09-744-313A-1 (1-465) x ABX50426 (1-305)

QY 180 GlnGluThrLeuGlnLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuA199  
Db 1 CAGGAATATCTGCAGAACTCTGCAACATCCAGATTGATGATCACTCTCGCA 60  
QY 200 AspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysLeuProAspVal 219  
Db 61 GATTTCTTCTCCCAATGTTGGGAAACACAGTTTCTGATAAGATCTACCAGACGTA 120  
QY 220 AsnLeuGlyLysLeuLeuLysSerValProGlyLysLeuMetLysGlyGlnHis 239  
Db 121 AATCTTGGAAATATATAAATCTGTTCTCGAAACATTAATGAAGAAAGTCAACAT 180  
QY 240 LeuGluProPheLeuMetAsnSerCysGluSerProLysProLysProSer 259  
Db 181 TTGAACCTTTTCATCATGAATTTTCAATTTCTTGTGAATCTCAAGCCTTAACCGAT 240  
QY 260 ArgProGluLeuThrLeuLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAsp 279  
Db 241 AAACAGACTGACCAATCTCAGGCTACATCAGAGATTAATACAGCTTTTATGAT 300  
QY 280 Leu 280  
Db 301 CTG 303

RESULT 12  
ABK63936  
ID ABK63936 standard; cDNA; 1900 BP.

AC ABK63936;

XX 18-JUN-2002 (first entry)

XX cDNA encoding human secreted protein sequence #21.

XX Human secreted protein; autoimmune disease; hyperproliferative disorder;  
KW cardiovascular disorder; cerebrovascular disorder; infection; cancer;  
KW nervous system disorder; ocular disorder; epithelial cell proliferation;  
KW wound healing; skin aging; sunburn; transplantation; chemotaxis;  
KW tissue regeneration; food additive; preservative; cytostatic; cardant;  
KW antiviral; antiallergic; antiinflammatory; antibacterial; antifungal;  
KW gene; ss.

OS Homo sapiens.

XX W0200216388-A1.

XX PD 28-FEB-2002.

XX PF 17-JAN-2001; 2001WO-US01394.

XX PR 18-AUG-2000; 2000US-226278P.

XX PR 05-JAN-2001; 2001US-259805P.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Rosen CA, Komatsoulis GA, Baker KP, Birse CE, Soppet DR, Olsen HS;

PI Moore PA, Wei P, Ebner R, Duan DR, Shi Y, Choi GH, Fischella M;

PI Ni J;

XX WPI; 2002-292053/33.

DR P-FSDB; AAU91368.

XX Isolated nucleic acids encoding 21 secreted proteins useful for the  
PT diagnosis and treatment of e.g. cancer, HIV infection, stroke and  
PT rheumatoid arthritis -

XX Claim 1; Page 451; 549pp; English.

XX The present invention relates to the isolation of novel human secreted  
CC proteins, and the polynucleotide sequences encoding them. The secreted  
CC proteins are useful to prevent, treat or ameliorate a medical condition  
CC in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or  
CC sheep. The secreted proteins are also useful in diagnosing a pathological  
CC condition or susceptibility to a pathological condition. Antibodies to  
CC the secreted proteins can also be used in alleviating symptoms associated  
CC with disorders and in diagnostic immunoassays e.g. radioimmunoassays or  
CC enzyme linked immunosorbent assays (ELISA). Disorders which can be  
CC diagnosed or treated include autoimmune diseases e.g. rheumatoid  
CC arthritis, hyperproliferative disorders e.g. cancer, cardiovascular  
CC disorders e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral  
CC ischaemia, angiogenesis, nervous system disorders e.g. Parkinson's  
CC disease, infections caused by bacteria, viruses and fungi and  
CC ocular disorders e.g. corneal infection. The polypeptides can also be  
CC used to aid wound healing and epithelial cell proliferation, to prevent  
CC skin aging due to sunburn, to maintain organs before transplantation, for  
CC supporting cell culture of primary tissues, to regenerate tissues and in  
CC chemotaxis. The polypeptides can also be used as a food additive or  
CC preservative to increase or decrease storage capabilities.  
CC ABK63916-ABK63965 encode human secreted protein sequences.

XX SQ Sequence 1900 BP; 657 A; 328 C; 400 G; 512 T; 3 other;

Alignment Scores:

Pred. No.: 5,44e-16 Length: 1900  
Score: 262.00 Matches: 98  
Percent Similarity: 42.65% Conservations: 79  
Best Local Similarity: 23.61% Mismatches: 154  
Query Match: 10.75% Indels: 84  
DB: 24 Gaps: 15

US-09-744-313A-1 (1-465) x ABK63936 (1-1900)

QY 56 IleGluGluGlyIleValValMetGlu-----AspAspSerProValGluAlaVal 72

Db 277 TTGAAGCATGAATTAATCTTAATAGAAAGACGACAGACCTTCAGCTGCACATGGCA 336

QY 73 SerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIleProTyrValAsp 92

Db 337 AGAACCGATTGCTGGTGTGTAACAACTTCGTCATGTGGAAGCCTTCATCACAGTGGAGAG 396

QY 93 PhePheGluAspProSerSerGluArgLysGluLysLysGluArgIleProValPheCys 112

Db 397 GTTACAGAA-----GAGATGTGTGACCAATTCGCTACTTT 435

QY 113 IleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSerVal 132



RESULT 13  
AAI61094

RESULT 13  
AAI61094





CC The proteins are useful for identifying compounds that modulate their  
 CC activity and production. The sequences of the invention are  
 CC useful for treating diseases such as hyperproliferative disorders  
 CC (e.g. cancer), immune deficiency disorders (e.g. severe combined  
 CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple  
 CC sclerosis), blood disorders (e.g. thrombocytopenia), inflammatory  
 CC disorders (e.g. arthritis), infectious disorders (e.g. hepatitis),  
 CC allergic conditions (e.g. asthma), neurodegenerative disorders (e.g.  
 CC Alzheimer's disease), liver fibrosis, coagulation disorders (e.g.  
 CC haemophilia), and tumours. The polynucleotide sequences of the  
 CC invention are also useful in gene therapy. ABK35610-ABK36232 represent  
 CC the cDNA sequences of the invention that encode for novel human  
 CC secreted proteins.

XX Sequence 2494 BP; 845 A; 445 C; 547 G; 657 T; 0 other;

# Alignment Scores:

Pred. No.: 1-39e-14 Length: 2494  
 Score: 249.50 Matches: 100  
 Percent Similarity: 42.75% Conservative: 77  
 Best Local Similarity: 24.15% Mismatches: 155  
 Query Match: 10.24% Indels: 83  
 DB: 24 Gaps: 14

US-09-744-313a-1 (1-465) x ABK35847 (1-2494)

Qy 56 IleGluGluGlyLeuValValMetGlu-----AspAspSerProValGluAlaVal 72  
 Db 908 TTGAAGGATGAAATATCTTAATAGAAAGACGACACAGCTTCAGCTGCACATGGCA 967  
 Qy 73 SerThrProAsnThrProArgAsnLeuAlaAlaTPlsyleSerileProTyrValAsp 92  
 Db 968 AGAAGCGATGGTGGTGTGAACCTTGGCATGTGGAAGCCCTCATCACCGATGGAGAG 1027  
 Qy 93 PhePheGluAspProSerSerGluArgLysGluLysGluArgLysGluArgLysGlu 112  
 Db 1028 GTTACAGAA-----GAGATGGTGAGCAATGCCATGTTACTTT 1066  
 Qy 113 IleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrpSerVal 132  
 Db 1067 GTCATGTGAAGCCTTACAAGAA-----GTTGGAGGAGTTCGAACTAAGAACTGGACGGTC 1120  
 Qy 133 TyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAla 152  
 Db 1121 CCCAGAAGGCTCAGCGAGTTTCAGATTATACCGGAAACTCAGTGAGTGGCGTCCCTTCT 1180  
 Qy 153 PheProAspAlaGlnLeuProSerLysArgLysGluLeuGlyProLysAsnTyrGlu----- 170  
 Db 1181 TTAATAAAGTCCAGTGGCTTCTCTTAGCAAGCTGCCTTTCARATCTATAGATCAAAG 1240  
 Qy 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGluLysLeuGlnHisPro 190  
 Db 1241 TTTATGGAAGTGAAGATCAATTAATAAGTTTTACAGAATCTGCTTTCAGATGAA 1300  
 Qy 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
 Db 1301 AGACTGTGTGACAGTGAAGCACTTTATGCTCTTCTGAGCCCTTCTCCTGACTACCTCAAG 1360  
 Qy 211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLeuLysSerValProGly 230  
 Db 1361 GTTATCGAC-----GTGCGAGGG 1378  
 Qy 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSer 250  
 Db 1379 -----AAAAAATCTTTTTCATTCCTCATTTTG----- 1411  
 Qy 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
 Db 1411 ----- 1411  
 Qy 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290  
 Db 1412 -----GAAAGACTTCTCGGACTTCTTC-----TCCACCAGGAGGAG 1453

Qy 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
 Db 1454 ACAGAGGAGGACAGTACCTGTGCAGATTATGGTGAT-----CATGTGATGGAGGAAA 1507  
 Qy 311 AspTyrLeu-----MetTyrValGlyArgValValPheGlnValPro--- 324  
 Db 1508 GACGCTTGGCTGAACCATGTTTCATGTTGATTGGGGAG---ATTTTGAACCTTCAGGA 1564  
 Qy 325 -----AspTyrLeuHisHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeu 342  
 Db 1565 ATGTTTAAATGGGGAGAGAACATAATATGCCC-----CTGTTTCAGGTCACTTTG 1615  
 Qy 343 GluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArg 362  
 Db 1616 GAAGA-ACCATCAACAACAAATCCGGGACACAGTCAGCTGGATTTTCAGTGAGCAATG 1674  
 Qy 363 LeuValSerLeuIleThrLeuLeuArgAspAlaIlePhe-----CysGlu 377  
 Db 1675 TTGGTTTACTACATCAATATTTCCGGGATGCTTTTGGCCAAATGGGAGTTGGCACCA 1734  
 Qy 378 AsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGlu 397  
 Db 1735 CCGACCAATCAGAGCAAGAGCAAAAGTCAGAAACAAACAGAGAGCAGACGAAAG 1794  
 Qy 398 MetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGlu 417  
 Db 1795 CTGCTTGAACAACATTCAGATATGCTTCAGAGCTTGTTCAGACAGCAAAATGCCGCCAC 1854  
 Qy 418 SerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyr 437  
 Db 1855 GGTATATAAATAATTCATGCTCAAGAAACAAAGACCAACCAAGCATCTGTATAT 1914  
 Qy 438 ValLeuLeuAspIleValIleGlnGluLeuPheProGluLeu 451  
 Db 1915 CGCGTGAATGGAACCTGCTGCTAATTGAACCTGTCTCTGAGCTG 1956

Search completed: January 31, 2004, 02:52:51

Job time : 403 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 02:43:52 ; Search time 2927 Seconds  
(without alignments)  
3861.148 Million cell updates/sec

Title: US-09-744-313A-1  
Perfect score: 2437  
Sequence: 1 MYLHCLIFRNTQKRGSEF.....ELFPELNKVKQEVTSVTSWM 465

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+ p2n.model -DEV=xlh  
-Q=cgnt\_1/USPTO/US09744313/runat 30012004 113535 4196/app\_query.fasta\_1.647  
-DB=EST -QFMT=fastap -SUFFIX=rest -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-USER=US09744313 @CGN 1 1 2810 @runat 30012004 113535 4196 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.\*  
1: em\_estba.\*  
2: em\_esthum.\*  
3: em\_estin.\*  
4: em\_estnu.\*  
5: em\_estov.\*  
6: em\_estpl.\*  
7: em\_estro.\*  
8: em\_etc.\*  
9: gb\_est1.\*  
10: gb\_est2.\*  
11: gb\_hc.\*  
12: gb\_est3.\*  
13: gb\_est4.\*  
14: gb\_est5.\*  
15: em\_estfun.\*  
16: em\_estom.\*  
17: em\_gss\_hum.\*  
18: em\_gss\_inv.\*  
19: em\_gss\_pln.\*  
20: em\_gss\_vrt.\*  
21: em\_gss\_fun.\*  
22: em\_gss\_man.\*  
23: em\_gss\_mus.\*  
24: em\_gss\_pro.\*  
25: em\_gss\_rod.\*  
26: em\_gss\_pig.\*  
27: em\_gss\_vrl.\*  
28: gb\_gss1.\*

29: gb\_gss2.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the total being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2347	96.3	3064	11	AK049371	Mus muscu
2	1949	80.0	2634	11	AK082389	Mus muscu
3	1517	62.2	917	13	BX372281	BX372281
4	1404.5	57.6	905	14	CB182722	CB182722
5	1348.5	55.3	1043	9	AL574298	AL574298
6	1345	55.2	994	13	BUI18410	BUI18410
7	1271	52.2	818	12	BQ191447	BQ191447
8	1266	51.9	1096	12	BQ053127	BQ053127
9	1230.5	50.5	921	13	BQ230190	BQ230190
10	1220	50.1	903	9	AL578668	AL578668
11	1190.5	48.9	915	13	BUI243572	BUI243572
12	1190	48.8	1105	10	BG256631	BG256631
13	1172	48.1	1093	10	BG293966	BG293966
14	1162	47.7	754	13	BUI73460	BUI73460
15	1151.5	47.3	1076	10	BH874550	BH874550
16	1148	47.1	664	14	CB136167	CB136167
17	1127	46.2	1096	14	CD500048	CD500048
18	1117	45.8	656	10	BB386253	BB386253
19	1111	45.6	722	9	AJ393454	AJ393454
20	1084	44.5	625	10	BG536529	BG536529
21	1083	44.4	643	14	CB114674	CB114674
22	1069.5	43.9	638	10	AW957704	AW957704
23	1059.5	43.9	726	13	BX303054	BX303054
24	1064	43.7	952	13	BQ171610	BQ171610
25	1059.5	43.5	810	10	BG191057	BG191057
26	1059	43.5	608	12	BW222250	BW222250
27	1059	43.5	752	13	BUI282728	BUI282728
28	1056	43.3	608	12	BM829345	BM829345
29	1053	43.2	616	14	CB054303	CB054303
30	1045	42.9	705	13	BX17013	BX17013
31	1045	42.9	936	10	BG026479	BG026479
32	1044	42.8	604	14	CB130920	CB130920
33	1042	42.8	802	14	CA431124	CA431124
34	1038	42.6	606	12	BM699448	BM699448
35	1025.5	42.1	861	10	BG214306	BG214306
36	1021	41.9	817	10	BF692296	BF692296
37	1020	41.9	897	10	BG180026	BG180026
38	995	40.8	828	13	BUI136331	BUI136331
39	988.5	40.6	936	13	BQ336152	BQ336152
40	979.5	40.2	800	10	BF581597	BF581597
41	976	40.0	656	9	AL957064	AL957064
42	972	39.9	675	10	BG410145	BG410145
43	968	39.7	558	14	CB160118	CB160118
44	963	39.5	923	9	AL561693	AL561693
45	962	39.5	639	13	BUI333500	BUI333500

#### ALIGNMENTS

RESULT 1  
AK049371

LOCUS  
DEFINITION

AK049371 3064 bp mRNA linear HTC 05-DEC-2002  
Mus musculus ES cells cDNA, RIKEN full-length enriched library,  
clone: C330035N22 product: HYPOTHETICAL 102.9 KDA PROTEIN (SORTING  
NEXTIN 14) homolog [Homo sapiens], full insert sequence.

ACCESSION  
AK049371.1

VERSION  
GI:26340105

KEYWORDS  
HTC; CAP trapper.

SOURCE  
Mus musculus (house mouse)

ORGANISM  
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Muridae; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE AUTHORS TITLE JOURNAL MEDLINE PUBMED	1	Carninci, P. and Hayashizaki, Y. High-efficiency full-length cDNA cloning Meth. Enzymol. 303, 19-44 (1999) 99279253 10349636
	2	Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes Genome Res. 10 (10), 1617-1630 (2000) 20499374 11042159
	3	Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Katsunai, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Ogawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer Genome Res. 10 (11), 1757-1771 (2000) 20530913 11076861
	4	Kawai, J., Shingawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamakawa, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H., Kueh, P., Lewis, S., Matsuo, Y., Nikaide, I., Pesole, G., Quackenbush, J., Schriml, L.M., Stauber, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N., Carninci, P., de Bonaldo, M.F., Brownstein, M.J., Bult, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D., Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombardi, P., Nordone, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H., Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H., Toyooka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L., Wynshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawaji, H., Kohsaki, S. and Hayashizaki, Y. Functional annotation of a full-length mouse cDNA collection Nature 409 (6821), 685-690 (2001) 21085660 11217851
REFERENCE AUTHORS TITLE JOURNAL MEDLINE PUBMED	5	The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs Nature 420, 563-573 (2002) 6 (bases 1 to 3064)
	6	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, M., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T., Katoch, H., Kawai, Y., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N., Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Shibata, K., Shingawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, P., Takaku-Akahira, S., Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome
	7	US-09-744-313A-1 (1-465) x AK049371 (1-3064)
	QY	11 ArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerIleIleIys 30
REFERENCE AUTHORS TITLE JOURNAL MEDLINE PUBMED	Db	1570 AGAACACACACAAAGAGGAGATCGTTTGGATCAGCAGATAGTAGCAATATAA 1629
	QY	31 GlyValPheLysSerThrThrMetGluCiyAlaMetLeuProAsnTyrGlyValalaGlu 50
	Db	1630 GCGCGTATTCAAGAGTACGACAAATGAGGAGCTGTGCTGCTTAATTCGGGGTGCTGAA 1689
REFERENCE AUTHORS TITLE JOURNAL MEDLINE PUBMED	8	Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-res@gscc-riken.go.jp, URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222, Fax: 81-45-503-9216)
	9	cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN, Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site for further details. URL: http://genome.gsc.riken.go.jp/ URL: http://fantom.gsc.riken.go.jp/.
	10	Location/Qualifiers 1. 3064 /organism="Mus musculus" /mol_type="mRNA" /strain="C57BL/6J" /db_xref="PANTOM:DB:C330035N22" /db_xref="taxon:10090" /clone="C330035N22" /cell_type="ES cells" /clone_lib="RIKEN full-length enriched mouse cDNA library" 124. 2937 /note="unnamed protein product; HYPOTHETICAL 102.9 KDA PROTEIN (SORTING NEXIN 14) homolog (Homo sapiens) (SPT1Q9BSD1, evidence: PASTY, 93.3%ID, 99.6%length, match=2781) putative" /codon_start=1 /protein_id="BAC33716.1" /db_xref="GI:126340106" /translations="MGSWRTICGRLLKQRLDVGREICRQYLPFCFLLCVLSVAIL LNRYLHVMFWSGVAVVTFYCSGLPDSLLNFIFFTKYKQLGQLPQGHSCA VCGKVKCHRPSPILLNYPWLDLVSSKVDASLSEVLELVLENVYFWYRVDODE SFVDELRTIRPFASVLRIRHKVDIPSIITKLLKAAKHIEVIKAROKVNTYVL QQALBEYGPBLHVALNRDELOYLKLTLELPPILPKATDCSUTLLIREILSG SVLPSSDLFLADFTVPHLLIIIDDSPPPEKAPSPVFLQKFAEPKPKPSVLIK LEKIQEQQLLFRFMNFKQGVAVHLOFCLTVEFNDRILRPLSNDMLSHBE LOKITYCLDSEIDKIRPDFIVIEIQRVLAEGYIDVVKLOTWRCIFEAYEHLVSL ENVTMPCHSDRYFROLLGASPTNSKFNSTOKRSGSGISGSKIQGVKFSST TEGAVLPKNGVAGEDDFIEGIVMEDDSPEVAVSTNPTNLAWKISIPVDFF EDSSERKKKKRIPVFCIDVERNDRAVGHPEHWSVTRYLYEFLVLESKLFEFGT FPDALPSKRIIGPKNYEFLKSEKFEQKLQVLPSELSOLLALFSLPNGETQ FDLKLPSNGLIKISVPGKLMKKEQHLPEFIMSPINCSGPKFSPHITLSP TSENKLLNDLPKNANRAENTKQNYFMVMTVDGVYDLYLVGVGVQVDPDM LHLLMGRILFKNTLEMDYDYLQCKLEQLQFQEHRLVSLITLRDAICENTEPRL QDKQKGAQTFEEMNTIIPDLIVKIGETKIESIRLLFDLQQLQPVLNKQLTYVLDDI VIQELPPELNKVKQKATSMTSWM" 3044. 3049 /note="putative" 3064 /note="putative"
	11	polyA_signal 886 a 637 c 704 g 837 t polyA_site BASE COUNT 886 a 637 c 704 g 837 t ORIGIN Alignment Scores: Pred. No.: 7,596-245 Length: 3064 Score: 2347.00 Matches: 446 Percent Similarity: 99.56% Conservative: 7 Best Local Similarity: 98.02% Mismatches: 2 Query Match: 96.31% Indels: 0 DB: 11 Gaps: 0



Qy 51 GlyGluAspPheIleGluGluGlyLeuValMetGluAspSerProValGlu 70  
Db 1690 GCGAAGATGACTTTATTGAAGAGGATTTGGTAATGAGGATGACTCTCCAGTAGAA 1749  
Qy 71 AlaValSerThrProAsnThrProArgSerLeuAlaAlaTrpLysIleSerIleProTyr 90  
Db 1750 GCTGTGAGCACACCTTCTCGAACTCTCTGAACTTGTCTGCTGAAATTAGCATTTCCGTAT 1809  
Qy 91 ValAspPheGluAspProSerGluArgLysGluLysLysLysGluArgIleProVal 110  
Db 1810 GTAGACTTTTGAAGATCCCTCTCTGAAAGAAAGAGAGAGAGAGATTCCTGTG 1869  
Qy 111 PheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrp 130  
Db 1870 TTTTGTATCCATGTTGAAAGAAACACACAGCAGCAGTGTGACAGCCTGAGCATTTG 1929  
Qy 131 SerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
Db 1930 TCTGCTATAGAGATATCTTGAATTTCTTGTGATCAAACTAACTAACAGATTTTCAT 1989  
Qy 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu 170  
Db 1990 GGCACATTTCTGTGCTGCTGCTTCTCAATCCAAAGATCATTTGGCCCAAAACTATGAG 2049  
Qy 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190  
Db 2050 TTTTGAAGTCGAGAGAGAGAGGTTCCAGGAGTATCTCGAAACTTTGTGCGACACCG 2109  
Qy 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
Db 2110 GAGCTGAGTAACAGCAGCTTCTGCTGCTTCTCTCCCCCAACGGTGGGAGACACAG 2169  
Qy 211 PheLeuAspLysIleLeuProAspValLeuLeuGlyLysIleLysSerValProGly 230  
Db 2170 TTTCTAGATAGATATCTTCCAGATGAATCTTGGAAATATTAAGATCTGTTCTCTGG 2239  
Qy 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSer 250  
Db 2230 AAATAATGAAGAGAGAGGTCAGATTTGGAGCATTCATCATGATTTTCAATATCT 2289  
Qy 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
Db 2290 TCGCAATCTCCCAACCTTAACCGAGTCGCGCAGAACTGACGATTTCTCAGCCCACTTCC 2349  
Qy 271 GluAsnAspLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290  
Db 2350 GAAACCAACAGAGAGCTTTTCAACGATCTGTTTAAGAATAATGCAAAACCGCGCTGAGA 2409  
Qy 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
Db 2410 ACAGAGCGAAAGCAAAATCAGAACTACTTTCATGGAGGTGATGCTGTGGAGGAGTCTAT 2469  
Qy 311 AspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHisLeu 330  
Db 2470 GATTACCTGATGATGAGGCGAGTAGTTTTTCAAGTCCCGAGATGCTTCATCATCTT 2529  
Qy 331 LeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
Db 2530 CTAATGGGAACCTCGAATCTCTTTAAGAACACCTTGAAGATGTACACAGACTACTACCTC 2589  
Qy 351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerIleLeuThrLeuLeu 370  
Db 2590 CAGTGAACCTGAGCAGCTGTTTTCAGGAGCACCGGCTGCTCTCCCTCATCACCTTCTC 2649  
Qy 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
Db 2650 AGAGATGCTATATTTTGGAAATAATGAACTGAACTGCTCTCTCCAGATGAGCAAAAGA 2709  
Qy 391 AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
Db 2710 GCAAAACAGACTTTTGAAGAAATGATGAATTACATTTCCAGATCTCTGATGATCAAGTATC 2769

Qy 411 GlyGluThrIleTyrTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnProVal 430  
Db 2770 GGTGAAGAACAACAGTACCAAGACATCACTCTCTGATGCTCTGACGAGCCAGTG 2829  
Qy 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnLeuPheProGlu 450  
Db 2830 CTCAACAGCAGCTGACCTATGTTGTTGCTGACATTTGATACAGAGGTTTTCAGAG 2889  
Qy 451 LeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
Db 2890 CTAATAAGGTACAAAGAGAGCTACCTCCATGACATCTCTGGATG 2934  
RESULT 2  
AK082389 2634 bp mRNA linear HTC 05-DEC-2002  
LOCUS AK082389  
DEFINITION Mus musculus 0 day neonate cerebellum cDNA, RIKEN full-length  
enriched library, clone:C230045H05 product:HYPOPHYSICAL 102.9 KDA  
PROTEIN (SORTING NEXIN 14) homolog [Homo sapiens], full insert  
sequence.  
ACCESSION AK082389  
VERSION AK082389.1 GI:26100628  
KEYWORDS HTC; CAP trapper.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
REFERENCE 1  
AUTHORS Carninci, P. and Hayashizaki, Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Meth. Enzymol. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2  
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)  
MEDLINE 20493374  
PUBMED 11042159  
REFERENCE 3  
AUTHORS Shibata, K., Itoh, M., Aizawa, K., Nagaoka, S., Sasaki, N., Carninci, P.,  
Konno, H., Akiyama, J., Nishi, K., Kitsumai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishine, T., Harada, A.,  
Yamanoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwaka, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE RIKEN integrated sequence analysis (RISA) system--384-format  
sequencing pipeline with 384 multicapillary sequencer  
JOURNAL Genome Res. 10 (11), 1757-1771 (2000)  
MEDLINE 20530913  
PUBMED 11076861  
REFERENCE 4  
AUTHORS Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y.,  
Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S.,  
Aizawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanaka, I.,  
Saito, T., Okazaki, Y., Gotobori, T., Bono, H., Kasukawa, T., Saito, R.,  
Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T.,  
Fleischmann, W., Gaasterland, T., Gissi, C., King, B., Kochiwa, H.,  
Kuehl, P., Lewis, S., Matsuo, Y., Nikaido, I., Pesole, G.,  
Quackenbush, J., Schriml, L.M., Staubli, F., Suzuki, R., Tomita, M.,  
Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Kono, H.,  
Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, N.,  
Carninci, P., de Bonaudo, M.F., Brownstein, M.J., Sult, C.,  
Fletcher, C., Fujita, M., Gariboldi, M., Gustincich, S., Hill, D.,  
Hofmann, M., Hume, D.A., Kamiya, M., Lee, N.H., Lyons, P.,  
Marchionni, L., Mashima, J., Mazzarelli, J., Mombaerts, P., Nordone, P.,  
Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,  
Sato, K., Schonbach, C., Seya, T., Shibata, Y., Storch, K.F., Suzuki, H.,  
Toyo-oka, K., Wang, K.H., Weitz, C., Whittaker, C., Wilming, L.,  
Wyshaw-Boris, A., Yoshida, K., Hasegawa, Y., Kawai, J., Kohseki, S.  
and Hayashizaki, Y.

```

TITLE      Functional annotation of a full-length mouse cDNA collection
JOURNAL    Nature 409 (6821), 685-690 (2001)
MEDLINE    21085660
PUBMED     11217851
REFERENCE  5
AUTHORS    The PANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE      Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL    Nature 420, 563-573 (2002)
REFERENCE  6 (bases 1 to 2634)
AUTHORS    Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Bono, H., Carninci, P.,
            Fukuda, S., Furuno, M., Hanagaki, T., Hara, A., Hashizume, W.,
            Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hirozane, T.,
            Hori, F., Inotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kasukawa, T.,
            Katoh, H., Kawaji, C., Kojima, Y., Kondo, S., Kono, H., Kouda, M.,
            Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
            Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Ohno, M., Ohsato, N.,
            Okazaki, Y., Saito, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N.,
            Sano, H., Sasaki, D., Shibata, K., Shinagawa, A., Shiraki, T.,
            Sogabe, Y., Tagami, M., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,
            Takeda, Y., Tanaka, T., Tomaru, A., Toya, T., Yasunishi, A.,
            Muramatsu, M. and Hayashizaki, Y.
TITLE      Direct Submission
JOURNAL    Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail: genome-res@gsr.riken.go.jp,
            URL: http://genome.gsc.riken.go.jp/, Tel: 81-45-503-9222,
            Fax: 81-45-503-9216)
COMMENT    cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Sciences Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL: http://genome.gsc.riken.go.jp/
            URL: http://fantom.gsc.riken.go.jp/
FEATURES   Location/Qualifiers
            1..2634
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="FANTOM_DB:C230045H05"
                /db_xref="taxon:10090"
                /clone="C230045H05"
                /tissue_type="cerebellum"
                /clone_lib="RIKEN full-length enriched mouse cDNA library"
                /dev_stage="0 day neonate"
            misc_feature
                1..2634
                    /note="HYPOTHEtical 102.9 KDA PROTEIN (SORTING NEXIN 14)
                    homolog [Homo sapiens] (SPTR|Q9BSD1, evidence: FASTV,
                    93.3%ID, 99.6%length, match=2781)"
BASE COUNT      720 a 570 c 624 g 720 t
ORIGIN
Alignment Scores:
Pred. No.:      1,876-201      Length:      2634
Score:          1949.00      Matches:      369
Percent Similarity: 99.73%      Conservative: 5
Best Local Similarity: 78.40%      Mismatches: 1
Query Match:     79.98%      Indels: 0
DB:              11          Gaps: 0
US-09-744-313A-1 (1-465) x AK082389 (1-2634)
Qy          10 PheArgAsnThrGlnIysArgGlyGluSerPheGlySerArgIleGlySerIlystle 29
Db          1509 TTCGGAAGCACACAAAAAGAGGAGGAGTCTTTGGGAATCAGCAGATAGTAGCAAAATT 1568
Qy          30 LysGlyValPheLysSerThrMetGluGlyAlaMetLeuProAsnThrGlyValala 49

```

```

DEFINITION BX372281 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
Homo sapiens cDNA clone CS0DL007YE10 3-PRIME, mRNA sequence.
ACCESSION BX372281
VERSION BX372281.1 GI:30441965
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 917)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished
CONTACT: Genoscope
Genoscope - Centre National de Sequencage
BP 191 9106 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3600.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0BA1044ZB11_CS04210_1&cluster=3600.r.
Contact : Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0BA1044ZB11_CS04210_1.
Paradise Avenue Genoscope sequence ID : CS0BA1044ZB11_CS04210_1.
FEATURES
source
1..917
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DL007YE10"
/cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED"
/cell_line="RAMOS CELL LINE"
/clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT 257 a 168 c 177 g 310 t 5 others
ORIGIN
US-09-744-313A-1 (1-465) x BX372281 (1-917)

Alignment Scores:
Pred. No.: 6,6e-155 Length: 917
Score: 1517.00 Matches: 293
Percent Similarity: 96.72% Conservative: 2
Best Local Similarity: 96.07% Mismatches: 9
Query Match: 62.25% Indels: 1
DB: 13 Gaps: 0

US-09-744-313A-1 (1-465) x BX372281 (1-917)

QY 135 ArgTyrLeuGluPheThrValLeuGluSerLysLeuThrGluPheHisGlyValAlaPhePro 154
Db 915 AGATTCTTGAATCTTATGCTAGTGAATCAAACTAACAGAAATTCATGGTGCTATTCCT 856
QY 155 AspAlaGlnLeuProSerLysArgLeuLeuGlyProLysAsnTyrGluPheLeuLysSer 174
Db 855 GATGCCAGCTCCCTCTAGAGGATCATGGCCCAAAAATATGATATCTTAAGTCAA 796
QY 175 LysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHisProGluLeuSerAsn 194
Db 795 GNNAGGAAGAGTTCACAGATATCTACAGAAATCTTCGACATCCAGAACTGAGTAAT 736
QY 195 SerGlnLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLys 214
Db 735 AGTCAACTCTGTCAGACTTCTTCCCTTAATGGTGGGAAACACAAATTTCTTGTATAG 676
QY 215 IleLeuProAspValAsnLeuGlyLysIleLeuLysSerValProGlyLysLeuMetLys 234
Db 675 ATACTCCAGATGAATCTTGGGAAAATATANAATCTGTCTCCGGAAAATTAATGAAA 616
QY 235 GluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerPro 254

```

```

Db 615 GAGAAAGGTGAGCATTTTGAACCTTTTATCATGAATTTTCAATTAATCTTGTGAGTCTCA 556
QY 255 LysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLys 274
Db 555 AGCCTTAAACCAAGTAGACCAAGTCCATTCCTCAGCCCTACTTTCAGAAAACCAACAG 496
QY 275 LysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLys 294
Db 495 AGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTGCTGAAATACAGAGAGAAAG 436
QY 295 GlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMet 314
Db 435 CAAATCAGAAATTTATTTATGAGGTGATGATCTGTAGAGGAGTCTATCATTTACCTGATG 376
QY 315 TyrValGlyArgValValPheGlnValProAspTyrLeuHisIleLeuMetGlyThr 334
Db 375 TATGTAGGACGGGTAGTTTTCAGGTTCTGACTGCTTCATCATCTCTTAATGGAACT 316
QY 335 ArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeu 354
Db 315 CGAATCTCTTTAAAAACCCCTGGAAATGATATCTGATTAATCTTCACTGTTAACTA 256
QY 355 GluGlnPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIle 374
Db 255 GAACGCTATTTTCAGGAGCACCGTTTGGTCTCACTCAATCACTTCTCAGAGATGCTATA 196
QY 375 PheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThr 394
Db 195 TTCTGTGAAAAACACTGAACCTGCTCTCCAGATTAAGCAAAAGAGGACCAACAGACT 136
QY 395 PheGluGluMetMetAsnTyrIleProAspLeuValLysCysIleGlyGluGluThr 414
Db 135 TTTGAGAAATGATGAATTACATTCAGATCTGTAGTCAAGTGTATTGGTGAAGAAACN 76
QY 415 -LysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysG 434
Db 75 CAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCAACCACTACTCAACAGCA 16
QY 434 nLeuThrTyrVal 438
Db 15 GCGTACTTATGTT 3

RESULT 4
CB182722
LOCUS CB182722
DEFINITION AGENCOURT 11383118 NIH MGC 164 Mus musculus cDNA clone
INAGR:30243103 5', mRNA sequence.
ACCESSION CB182722
VERSION CB182722.1 GI:28180813
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 905)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
CONTACT: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. David Rowe and Dr. Mina
cDNA Library Preparation: Invitrogen Corp
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: NDA0313 row: n column: 08
High quality sequence stop: 706.
FEATURES
source
1..905
/organism="Mus musculus"

```

```

/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:30243103"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 164"
Note: Vector: pCMV-SPORT6.1.cdb; Site 1: EcoRV; Site 2:
NotI; Non-normalized full-length enriched library from
pooled mouse embryonic limb, maxilla and mandible, day
10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
Cloned directionally, priming method: Oligo-dT. cDNA
enrichment: >1k bp, Average insert size 1.8k bp. Priming
sequence: 5' GACTGATCTGATCGGAGCGCCGCC(T) 3'. Tissue
contributed by, David Rowe. Library constructed by ResGen,
Invitrogen Corp."
BASE COUNT 276 a 207 c 194 g 226 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 1-27e-142 Length: 905
Score: 1404.50 Matches: 277
Percent Similarity: 95.61% Conservative: 6
Best Local Similarity: 93.58% Mismatches: 9
Query Match: 57.63% Indels: 4
DB: 14 Gaps: 1
US-09-744-313A-1 (1-465) x CB182722 (1-905)
Qy 143 GluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSerLysArg 162
Db 12 GAATCAAACTAAACAGATTTCATGTCACATTTCTCTGATGCTCAGCTCCATCCAAAGG 71
Qy 163 IleIleGlyProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGlnGluTyr 182
Db 72 ATCATTTGGCCCCAAAACATGATGTTTTGAATCGAAGAGAGAGAGATTCAGGAGAT 131
Qy 183 LeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeu 202
Db 132 CTGCAGAAATTTGTGCGAGCACCGGAGCTGAGTAAACAGCAGCTTCTGCTGACTTCTC 191
Qy 203 SerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGly 222
Db 192 TCCCCCAACGGTGGGAGACACAGTTTCTAGATAAGATCTCCAGATGTAAATCTTGA 251
Qy 223 LysIleIleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeuGluPro 242
Db 252 AAAATTATAAGTCTGTCTCTCGGAAACTAATGAAAGAGAAAGGTCCAGATTTCGAGCCA 311
Qy 243 PheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArgProGlu 262
Db 312 TTCATCATGAGTTTCATTATCTTTGCGAATCTCCCAACCTAAACCGAGTCGCGCAGAA 371
Qy 263 LeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLys 282
Db 372 CTGACGATTTCTAGCCCCCATCTCCGAAAACACACAGAGGTTTTCAACGATCTGTTTAA 431
Qy 283 AsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGlu 302
Db 432 AATAATGCAACCGCGTGAAGACACAGAGCGAAAGCAAAATCAGAACTACTTCATGGAG 491
Qy 303 ValMetThrValGluGlyValTyrAspTyrIleuMetTyrValGlyArgValValPheGln 322
Db 492 GTGATGACTGTGGACGGAGTCTATGATTAATCTGATGTATGAGCGAGTAGTTTTCAA 551
Qy 323 ValProAspTyrPheLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeu 342
Db 552 GTCCAGACTGGCTTCATCATCTTCTAATGGAACTCGAATCTCTTTAAGAACACCCCTG 611
Qy 343 GluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArg 362
Db 612 GAAATGTACACAGACTATTACTCCAGTGAAGCTGGAGCAGCTGTGTTTCAGGAGCACCG 671
Qy 363 LeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArg 382

```

```

672 CTGGTCTCTCCCTCATCCCTTCTCAGAGATGCTATATTTTGTGGAAATCTAGAACCTCGC 731
383 SerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIle 402
732 TCTCTCCAGATAGCAAAAGAGGCAAAACAGACTTTTTCGAAATGATGATTAACAT 791
403 ProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArgLeu 421
792 CCAGATCTGATAGTCAAGTGTATCGGTGAAGAACCCAGTACGAAAGCTTCAGAACTC 851
422 LeuPheAspGlyLeu---GlnGlnProValLeuAsnLysGln 434
852 CTGTTTATGGCGCTCGCAGCAGCCACCTGCTTCTCAACAGCAG 893

RESULT 5
AL574298/c
LOCUS
DEFINITION AL574298 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CSODI057YD12 3-PRIME, mRNA sequence.
ACCESSION AL574298
VERSION AL574298.2 GI:31312617
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1043)
Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished
On Feb 16, 2001 this sequence version replaced gi:12934370.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3600.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CSODI057YD12&cluster=3600.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
Faraday Avenue Genoscope sequence ID : CSODI057YD12B06NP1.

FEATURES
source
1..1043
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CSODI057YD12"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(GT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."

BASE COUNT 307 a 196 c 182 g 338 t 20 others
ORIGIN
Alignment Scores:
Pred. No.: 2-04e-136 Length: 1043
Score: 1348.50 Matches: 273
Percent Similarity: 91.39% Conservative: 3
Best Local Similarity: 90.40% Mismatches: 23
Query Match: 55.33% Indels: 5
DB: 9 Gaps: 1
US-09-744-313A-1 (1-465) x AL574298 (1-1043)
Qy 165 GlyProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGln 184
Db 1041 GGCCCCAAATATGATGATTTT-----AAAKYCAARGGGGAGGTCACAGATATTCAMC 988
Qy 185 LysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerPro 204

```

```
...
987 AGAAGCTTCGCGCATCCAGACACRAGTAGTAAAGCA-CCTCTGGCAAC-TTCTTCCCGCT 930
QY 205 AAGGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIle 224
Db 929 AATGGTGGGGAACAATTTCTTGAAGATACCTCCAGATKAAATCTTGGGAATTT 870
QY 225 IleLysSerValProGlyLysLeuMetLysGlyLysGlyGlnHisLeuGluPro-PheIle 244
Db 869 AATAAATCTGTCCTGGAAACCTAATGAAGAGAGAAAGCTCAGCATTTTGGAACTTTTAT 810
QY 244 eMetAsnPhelLeuSerCysGluSerProLysProLysProSerArgProGluLeuTh 264
Db 809 CATGAATTTCAATAATCTTGTGAGTCTCCGAAGCTAAACCAAGTACAGCAGAACTGAC 750
QY 264 rIleLeuSerProThrSerGluAsnLysLysLeuPheAsnAspLeuPheLysAsnAs 284
Db 749 CATTTCTAGCCCTACTTTCAGAAAAACAAGAAAGCTTTTCAATGATCTGTTTAAAAATAA 690
QY 284 nAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMe 304
Db 689 TGCAACCGTGTGAATATACAGAGAGAAAGCAAAATCAGATTAATTTATGGAGGTGAT 630
QY 304 tThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValPr 324
Db 629 GACTGTAGAAGGAGCTATGATTAATCTGATGTATGTAGACGGGTAGTTTCCAGGTTC 570
QY 324 oAspTrpLeuHisLysLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMe 344
Db 569 TGACTGGGTTCATCATCTCTTAATGGAACTCGAATCTCTTTAAAAACACCCCTGGAAAT 510
QY 344 tTyrThrAspTyrTyrLeuGlnCysLysLeuGlnLeuPheGlnGluHisArgLeuVa 364
Db 509 GATTAATGATTAATCTCTTCACTGTAAGTAACTAGACAGCTATTTCCAGGACCCGTTGGT 450
QY 364 lSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLe 384
Db 449 CTCACCTATAACACTTCTCAGAGATGCTATATTCTGTGAAACACCTGACATMGATCTCT 390
QY 384 uGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAs 404
Db 389 CCAAGTAAGCAAAAAGGAGCAAAACAGACTTTTCAAGAAATGATGAATTAATTAATCCAGA 330
QY 404 pLeuLeuValLysCysIleGlyGluThrLysTyrGluSerIleArgLeuLeuPheAs 424
Db 329 TCTGTAGTCAAGTGTATTGGTGAAGAACCAAGTATGAAGCATCAGACTTCTGTGTTGA 270
QY 424 pGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIle 444
Db 269 TGGCTTACAGCAACCACTACTCAGACAGCTGACTTATGTTTATTGGACATTTGGAT 210
QY 444 eGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSerValThrSerTr 464
Db 209 ACAGGAACCTGTTTCCAGAGCTCAATAAGTACAAAGGAGGTTACCTCTGTGACATCTTG 150
QY 464 pMet 465
Db 149 GATG 146

RESULT 6
LOCUS BUI18410
DEFINITION 603144375P1 CSEQCHL16 Gallus gallus linear EST 25-NOV-2002
sequence.
ACCESSION BUI18410
VERSION BUI18410.1 GI:25326894
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 994)
```

## AUTHORS

Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, B.,  
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.  
A Comprehensive Collection of Chicken cDNAs  
Curr. Biol. 12 (22), 1965-1969 (2002)  
22335534  
12445392

## COMMENT

Contact: Simon Hubbard  
Department of Biomolecular Sciences  
University of Manchester Institute of Science and Technology (UMIST)

PO Box 88, Manchester, M60 1QD, UK

Tel: 01612008930

Fax: 01612360409

Email: Simon.Hubbard@umist.ac.uk.

## FEATURES

source

1..994  
Location/Qualifiers  
/organism="Gallus gallus"  
/mol\_type="mRNA"  
/strain="Compton line 151"  
/db\_xref="taxon:9031"  
/clone="CHEST141e19"  
/sex="Female"  
/tissue\_type="not cerebrum or cerebellum"  
/dev\_stage="adult"  
/lab\_host="DH10B"  
/clone\_lib="CSEQCHL16"  
/note="Organ: brain; Vector: pBluescript II KS(+); Site 1:  
EcoRI; Site 2: NotI; Modification of pBluescript II KS(+)  
[stratagene] vector to accommodate cDNA produced with the  
T-trimmed protocol (Construction of uni-directionally  
cloned cDNA libraries from messenger RNA for improved 3'  
end DNA sequencing by Glenn Fu, et al. U.S. Patent # 6,387  
,624). Cut pBluescript II KS(+) with NotI and EcoRI.  
Ligate in double stranded adaptor containing BspI and  
BamHI sites [5'ggcgcgtgcagccggatccgaaaaaag]  
[5'aattctttttcggatccggggtgcacgc]"

BASE COUNT 301 a 202 c 227 g 264 t

## ALIGNMENT SCORES:

Pred. No.: 4,59e-136 Length: 994  
Score: 1345.00 Matches: 278  
Percent Similarity: 88.36% Conservative: 18  
Best Local Similarity: 82.99% Mismatches: 29  
Query Match: 55.19% Indels: 11  
DB: 13 Gaps: 3

US-09-744-313A-1 (1-465) x BUI18410 (1-994)

QY 125 HisGluProGluHisTrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSer 144  
Db 3 TATGAACCTGAGCAGCTGGTCACTTACAGAGAGTATCTGGAATTTATGCTTGTAGTCA 62  
QY 145 LysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIle 164  
Db 63 AGCTAACAGAAATTCATGCTACTTTCTGATGCTTCAGCTTCCCTCAAGAGAAATCAT 122  
QY 165 GlyProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGlnGlyLeuGln 184  
Db 123 GGGCCCCAAGACTATGATGTTCTTAAATCCAAAGGAGAGTTCAGAGATCTACAG 182  
QY 185 LysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerPro 204  
Db 183 AAATCTCTGCAACATCCAGAGCTAAGTAACACCAACTTTTAGCTGACTTCTGCTCCCT 242  
QY 205 AsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIle 224  
Db 243 AATCGAGGAGAGACTCAGTTTCTTGATAAAATTTGCGCTGATGTGAATCTGGTGAAGATT 302  
QY 225 IleLysSerValProGlyLysLeuMetLysGlyLysGlyGlnHisLeuGluProPheIle 244  
Db 303 ATAAATCTGTTCCAGAAAGCTGATGAAGAGAGAGAGTTCAGCATTGAGGCCATTCATC 362

```

QY 245 MetAsnPhelLeuAsnSerCysGluSerProLysProLysProSerArgProGluLeuThr 264
Db 363 ATGAAATTTATCAATCTCGGAATCTCCAGACCTAAACCCAGCAGCCCTGAGCTTACG 422
QY 265 IleLeuSerProThrSerGluAsnAsnLysLysLysLeuPheAsnAspLeuPheLysAsnAsn 284
Db 423 ATTCTGAGTCCCACTCTGAGAACACACAGAGCTTTTCAATGATTTATTCACAGATAAT 482
QY 285 AlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMet 304
Db 483 GCAACCCGCTCTGAGAAATACAGAAAGACGACAAATCAGAACTATTCATGCGAATGATG 542
QY 305 ThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValPro 324
Db 543 ACTGTAGAGGAGTCTATGACTACTTATTGTATGTTGTAGAGTGCTCTCCACATCTCT 602
QY 325 AspTrpLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMet 344
Db 603 GACTGCTTCATCATCTCTGATGAGGAGAGAAATCTCTTCAAAATACATTCGGAATCG 662
QY 345 TyrThrAspTyrTyrLeuGlnCysLysLeuGlnLeuPheGlnGluHisArgLeuVal 364
Db 663 TACACGACTATTTACTTGCATTTATAGTTAGAACAGCTATGTGAGGAGCACCGTTTGT 722
QY 365 SerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeu 384
Db 723 TCTCTCATCACTGCTGAGAGATGCTGTGTCTGTGAGATCTGAACCCGC-TCGTCT 781
QY 385 GlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMet---MetAsnTyr-IlePr 403
Db 782 CAGGACAACGACAGAGAGCAAGCAAGACTTTTTCGAAAGAACTGATGAGATACCTCTCC 841
QY 403 AspLeuLeuValLysCysIleLeuGluGlu-ThrLysTyrGluSerIle-ArgLeu 422
Db 842 AGATTTAATAGGCAAGTGCATTTGGGGGAGAGAGCTAAATACGAAAGCTCCAGGCTTC 901
QY 422 euPhe-----AspGlyLeuGlnProValLeuAsnLysGlnLeuThr---TyrVal 439
Db 902 TGTTTGATGCTGCTGCAGCAGCCCGAGTGGCTTCAACGACAGTTGACTTAACGTGCT 961
QY 439 euLeuAspIleValIle-GlnGluLeuPhePro 449
Db 962 GTTGGGACATCGGATTCAGGAATCCTTCCA 994

RESULT 7
BG191447 818 bp mRNA linear EST 21-APR-2001
RST10541 Athersys RAGE Library Homo sapiens cDNA, mRNA sequence.
BG191447
BG191447.1 GI:13713134
EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
AUTHORS Harrington,J.J., Sherf,B., Rundlett,S., Jackson,P.D., Perry,R.,
Cain,S., Leventhal,C., Thornton,M., Ramachandran,R., Whittington,J.,
Lerner,L., Costanzo,D., McElligott,K., Booser,S., Mays,R., Smith
,E., Veloso,N., Klika,A., Hess,J., Cothren,K., Lo,K., Offenbacher
,J., Danzig,J. and Ducar,M.
TITLE Creation of genome-wide protein expression libraries using random
activation of gene expression
JOURNAL Nat. Biotechnol. 19 (5), 440-445 (2001)
MEDLINE 21227151
PUBMED 11329013
COMMENT Contact: Scott J. Cain
Athersys, Inc.
3201 Carnegie Ave, Cleveland, OH 44115, USA
Tel: 216 431 9900
Fax: 216 361 9596
Email: scain@athersys.com
High quality sequence stop: 428.

```

```

FEATURES
source
Location/Qualifiers
1..818
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_line="HT1080"
/clone_lib="Athersys RAGE Library"
/note="See 'Creation of Genome-wide Protein Expression
Libraries using Random Activation of Gene Expression',
Nature Biotechnology, in press. Note that even though the
cell type indicated is HT1080, since a random activation
method was used, these sequence tags are not necessarily
expressed in HT1080 under normal circumstances."
BASE COUNT 275 a 159 c 229 g
ORIGIN
Alignment Scores:
Pred. No.: 4,23e-128 Length: 818
Score: 1271.00 Matches: 259
Percent Similarity: 95.97% Conservative: 3
Best Local Similarity: 94.87% Mismatches: 11
Query Match: 52.15% Indels: 5
DB: 10 Gaps: 0
US-09-744-313A-1 (1-465) x BG191447 (1-818)
QY 124 GlyHisGluProGluHisTrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGlu 143
Db 3 GGACACGAGCCTGA-CATTGGTCTGCTATAGAGATATCTTGAAATCTATGATCTTGA 61
QY 144 SerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSerLysArgIle 163
Db 62 TCAAACTAACAGAAATTCATGTCATTTCTGATGCCAGCTTCCTCTTAAGAGATC 121
QY 164 IleGlyProLysAsnTyrGluPheLeuLysSerLysArgGluGluPheGlnIleuVal 183
Db 122 ATTGGCCCCAAAATTTATGAATTTCTTAAAGTCAAAGAGGAGAGTTCCAAGATATCTA 181
QY 184 GlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSer 203
Db 182 CAGAAACTTCTGCAGCATCCAGAACTGAGTAAATAGTCAACTTCTGCAGACTTCTTTTC 241
QY 204 ProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLys 223
Db 242 CCTAATGTGGGGAACACACAAATTTCTTGATAGATACTACCAGATGAATCTTGGAAA 301
QY 224 IleIleLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPhe 243
Db 302 ATTATAAAATCTGCTCTCGGAAAACCTAATCAAGAGAAAAGTCAAGCATTTTGGAACTTT 361
QY 244 IleMetAsnPhelLeuAsnSerCysGluSerProLysProLysProSerArgProGluLeu 263
Db 362 ATCATGAATTTCAATTAATTTCTGTGAGTCTCCAAAGCCTAAACCAAGTAGACAGAACTG 421
QY 264 ThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsn 283
Db 422 ACCATTTCTGACCCCTACTTCAGAAAACACACAGAGACTTTTCAATGATCTGTTTAAAT 481
QY 284 AsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluVal 303
Db 482 AATGCANACCGTCTGTAANATAACAGAGAGAAACAAATATTTATTTATGAGAGTGTG 541
QY 304 MetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnVal 323
Db 542 ATGACTGTAGAGGAGTCTATGATTACCTGATGTATGTAGGACGGGTAGTTTTCAGAGTT 601
QY 324 ProAspTrpLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGlu 343
Db 602 CCTGACTGGCT-CATCATCTCTTAAATGGGAACTCGAATCTCTTNTAAAAACACCCCTG 660
QY 344 MetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnIleuVal 363
Db 661 ATGTATACTGATTTACTATCTTCTGATGTAACTAGACAGCTATTTTCAGGAGCACCTTT- 719

```



Qy 364 ValSerLeuThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSer 383  
 Db 720 GTCTCACTATACACTTCTCAAGATGCTATATTCTGTGAACACACCTGGACCTCGCTCT 779  
 Qy 384 LeuGluAspGlnGlyGlyAlaIleGlnThrPheGlu 396  
 Db 780 CTCAGAAATAG-CAAAAAGGAGC-AAACAGACTTTNGAG 816

RESULT 8  
 BQ053127  
 LOCUS BQ053127 1096 bp mRNA linear EST 29-MAR-2002  
 DEFINITION AGENCOURT\_6821786 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5934711  
 5', mRNA sequence.  
 ACCESSION BQ053127  
 VERSION BQ053127.1 GI:19812467  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1096)  
 AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: Dr. Daniel McVicar, DBS/NCI  
 cDNA Library Preparation: Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCM2121 row: e column: 16  
 High quality sequence stop: 675.

## FEATURES

source Location/Qualifiers

1..1096  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5934711"  
 /tissue\_type="natural killer cells, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_106"  
 /notes="Organ: blood; Vector: pOTB7; Site: 1: XhoI; Site 2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGACAGAG(G). Library constructed by Ling Hong in the  
 Laboratory of Gerald M. Rubin (University of California,  
 Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and  
 Superscript II RT (Life Technologies). Note: this is a  
 NIH MGC Library."

BASE COUNT 355 a 193 c 226 g 305 t 17 others

## ORIGIN

## Alignment Scores:

Pred. No.: 2,26e-127 Length: 1096  
 Score: 1266.00 Matches: 243  
 Percent Similarity: 99.18% Conservative: 0  
 Best Local Similarity: 99.18% Mismatches: 2  
 Query Match: 51.95% Indels: 0  
 DB: 12 Gaps: 0

US-09-744-313A-1 (1-465) x BQ053127 (1-1096)

Qy 221 LeuGlyValIleLeuIleLysSerValProGlyLysLeuMetLysGluLysGlnHisLeu 240  
 Db 7 TTGGGGAATTTATAAATCTGTCTCTGGAATACTAATGAAGAGAAAGTCAGCATTTG 66  
 Qy 241 GluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArg 260  
 Db 67 GAACCTTTTATCATGAATTTTCATTAATCTTGTGAGTCTCCAAAGCCCTAAACCAAGTAGA 126

Qy 261 ProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu 280  
 Db 127 CCAGAACTGACCACTTCTCAGCCCTACTCTCAGAAAACAACAAGAGCTTTTCAATGATCTG 186  
 Qy 281 PheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnThrPhe 300  
 Db 187 TTTAAAATAATGCAACCGCTGCTGAAAATACAGAGAGAAAGCAAAATCAGAAATTTATTTT 246  
 Qy 301 MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal 320  
 Db 247 ATGGAGGTGATGACTGTAGAGGAGTCTATGATTACTGTGATGATGTAGGACGGGTAGTT 306  
 Qy 321 PheGlnValProAspTyrPheHisLeuMetGlyThrArgIleLeuPheLysAsn 340  
 Db 307 TTCCAGGTTCCTGACTGGCTTCATCATCTCTTAATGGAACTCGAATCCTCTTTAAAAAC 366  
 Qy 341 ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGlnLeuPheGlnGlu 360  
 Db 367 ACCCTGAAAATGTATATCTGATTACTATCTTCACTGTAAACTAGAACAGCTATTTCAGGAG 426  
 Qy 361 HisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGlu 380  
 Db 427 CACCGTTTGGTCTCACTCAATACACATCTCTCAGAGATGCTATATCTGTGAAAACACATGAA 486  
 Qy 381 ProArgSerLeuGlnAspLysGlnLysGlyValAlaLysGlnThrPheGluGluMetMetAsn 400  
 Db 487 CCTCGCTCTCTCCAGATAGCAAAAGAGGCAAAACAGACTTTTGAAGAAAATGATGAAT 546  
 Qy 401 TyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArg 420  
 Db 547 TACATTCAGATCTGTTTAGTCAAGTGATTGGTGAAGAAACCAAGTATGAAGACATCAGA 606  
 Qy 421 LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu 440  
 Db 607 CTTCCTGTTGATGGCTTACAGCAACCACTACTCAACAGCAGCTGACTTATGTTTATTG 666  
 Qy 441 AspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSer 460  
 Db 667 GACATTGTGATACAGGAACCTGTTCCAGACTCAATAGGTACAAAAGGGAAGTACCTCT 726  
 Qy 461 ValThrSerTyrMet 465  
 Db 727 GTGACATCTTGGATG 741

## RESULT 9

BQ230190  
 LOCUS BQ230190 921 bp mRNA linear EST 02-MAY-2002  
 DEFINITION AGENCOURT\_7567151 NIH\_MGC\_70 Homo sapiens cDNA clone IMAGE:6054428  
 5', mRNA sequence.

ACCESSION BQ230190

VERSION BQ230190.1 GI:20411590

KEYWORDS EST.

SOURCE Homo sapiens (human)

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 921)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLCM13313 row: a column: 21  
 High quality sequence stop: 639.  
 Location/Qualifiers

## FEATURES

```
source
1..921
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6054428"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 70"
/notes="Organ: pancreas; Vector: pCMV-Sport6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT      307 a 168 c 181 g 265 t
ORIGIN
Alignment Scores:
Pred. No.:      1,33e-123      Length:      921
Score:          1230.50      Matches:    240
Percent Similarity: 96.41%      Conservative: 2
Best Local Similarity: 95.62%      Mismatches: 2
Query Match:    50.49%      Indels:    7
DB:             13      Gaps:      2

US-09-744-313A-1 (1-465) x BQ230190 (1-921)
QY      221 LeuGlyValIleLeuValProGlyLysLeuMetLysGluLysGluHisLeu 240
DB      5 TTGGGGAAATTAATAAACTCTTCTGGAAACTTAATGAAGAGAGGTCAGCATTTG 64
QY      241 GluProPheIleMetAsnPhelIleAsnSerCysGluSerProLysProLysProSerArg 260
DB      65 GAACCTTTTATCATGAATTTTCATTAATCTTGAGTCTCCAAAGCCCTAAACCAAGTAGA 124
QY      261 ProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu 280
DB      125 CCAGAACTGACCATTTCTGACCTACTTCCAGAAAACCAAGAGCTTTTCAATGATCTG 184
QY      281 PheLysAsnAsnAlaAsnArgAlaGluAsnThrGluAsnThrGluAsnGluAsnThrPhe 300
DB      185 TTAAAAAATAATGCAACCGTCTGAAATATACAGAGAGAAACCAAAATCAGAAATTTATT 244
QY      301 MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal 320
DB      245 ATGAGGTGATGCTGTAGAGAGTCTATGATTACCTGATGTATGTAGGACGGTAGTT 304
QY      321 PheGlnValProAspTrpLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsn 340
DB      305 TTCCAGGTTCTGACTGGCTTCATCATCTTAATGGGAACCTCGAATCCTCTTTAAAAAC 364
QY      341 ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGlu 360
DB      365 ACCCTGGAATGTATATCTGATTAATCTATCTAGTGTAAACTAGAACAGCTATTTCCAGGAG 424
QY      361 HisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGlu 380
DB      425 CACGTTTGCTTCATCTCACTCAACCTTTCAGAGATGCTATATTCTGTGAAAACACTGNA 484
QY      381 ProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsn 400
DB      485 CCTCGCTCTCTCCAGATAGCAAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAAT 544
QY      401 TyrIleProAspLeuLeuValLysCysIleGlyGluThrLysTyrGluSerIleArg 420
DB      545 TACATTCAGATCTGTTAGTCAAGTGTATTGGTGAAGAAACCAAGTATGAAAGCATCAGA 604
QY      421 LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu 440
DB      605 CTTCGTGTTGATGGCTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTTTATTG 664
QY      441 AspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGlu----- 457
DB      665 GACATTGTGATACAGGAACTGTTCCAGAGCTCCATAGGTACAAAAGGAAGTTAACT 724
```

```
QY      458 ---ValThrSerVal-----ThrSerTrp 464
DB      725 CTGGTGACATCTTTGGGATGTAAACACACCTTGG 757

RESULT 10
AL578668/c      903 bp      mRNA      linear      EST 01-JUN-2003
LOCUS      AL578668 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION      cDNA clone CS0DK004YA04 3-PRIME, mRNA sequence.
ACCESSION      AL578668
VERSION      AL578668.1 GI:12942962
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 903)
AUTHORS      Li, W.B., Gruber, C., Jesse, J. and Polayes, D.
TITLE      Full-length cDNA libraries and normalization
JOURNAL      Unpublished
COMMENT      Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 131 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 3600.r For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgi-bin/cluster.cgi?seq=CS0DK004BA02NP1&cluster=3600.r. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CS0DK004BA02NP1.

FEATURES             source
1..903
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK004YA04"
/cell_type="HELA"
/cell_line="HELA"
/clone_lib="Homo sapiens HELA CELLS COT 25-NORMALIZED"
/notes="1st strand cDNA was primed with a NotI-oligo(dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT      272 a 160 c 159 g 290 t 22 others
ORIGIN
Alignment Scores:
Pred. No.:      1.82e-122      Length:      903
Score:          1220.00      Matches:    239
Percent Similarity: 94.49%      Conservative: 1
Best Local Similarity: 94.09%      Mismatches: 14
Query Match:    50.06%      Indels:    1
DB:             9      Gaps:      0

US-09-744-313A-1 (1-465) x AL578668 (1-903)
QY      212 LeuAspLysIleLeuProAspValAsnLeuGlyLysIleLeuLysSerValProGlyLys 231
DB      902 CTTGATTAAGTACTTACCAGATGTAAATCTTGGAAAATATAAAATCTGTTCTCGAAA 843
QY      232 LeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPhelIleAsnSerCys 251
DB      842 CTAATGAAGAGAGAGGTCAGCATTTTGAACCTTTTATCATGATTAATTTCTGT 793
QY      252 GluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSerGlu 271
DB      782 GAGTCTCAAGCCTTAACCAAGTAGACCAAGACTGACCTTTCTAGCCCTTCTAGCCCTTCTAGAA 723
QY      272 AsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThr 291
DB      722 AACACAGAGAGGCTTTTCATGATCTGTTTAAAAAATAATGCAACCGTCTGGAATACA 663
```



QY 362 rgLeuValSerLeuThrLeuLeuArgAspAlaile-PheCysGluAsnThrGluPro 381  
Db 620 CGTAGTTCCTCTCAACACCTGCTGAGAGATGCTGTGTTCTGTGGAATCTGAACCG 679  
QY 382 ArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyr 401  
Db 680 CGCTCTGTCCAGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATAC 739  
QY 402 IleProAspLeuValLysCysIleGlyGluGlu-----ThrLysTyrGluSer 418  
Db 740 ATCCAGATTATAGGCAAGTGCCTGGGGAAGAGAGGCTAACTACGAAG-----GGC 793  
QY 419 IleArgLeuLeuPheAsp---GlyLeuGlnGlnPro---ValLeuAsnLys-GlnLeuTh 436  
Db 794 ATCAGGCTTCTAGTTCGATGAGCATGCGAGCAACCCAGATGCTCAAAACCCGTTAAC 853  
QY 436 rTyrValLeuLeuAspIle-ValIleGlnGlu-LeuPheProGlu 450  
Db 854 TTACGCTGCTTGGACATTCGGGANTTCAGAAACTCTTTCCAGAA 898

RESULT 12  
BG256631 1105 bp mRNA linear EST 13-FEB-2001  
LOCUS 602370809F1 NIH\_MGC\_92 Homo sapiens cDNA clone IMAGE:4478858 5',  
DEFINITION mRNA sequence.  
ACCESSION BG256631  
VERSION BG256631.1 GI:12766447  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 1105)  
TITLE NTH-MGC <http://mgc.nci.nih.gov/>.  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-x@mail.nih.gov](mailto:cgapbs-x@mail.nih.gov)  
Tissue Procurement: ATCC  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLAM10310 Row: a Column: 03  
High quality sequence stop: 645.  
Location/Qualifiers  
1. .1105  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4478858"  
/tissue\_type="embryonal carcinoma, cell line"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="NIH\_MGC\_92"  
/note="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;  
Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
Average insert size 2.5 kb. Library enriched for  
full-length clones and constructed by Life Technologies.  
Note: this is a NIH MGC Library."  
BASE COUNT 380 a 235 c 208 g 282 t  
ORIGIN

Alignment Scores:  
Pred. No.: 4.6e-119 Length: 1105  
Score: 1190.00 Matches: 237  
Percent Similarity: 96.00% Conservative: 3  
Best Local Similarity: 94.80% Mismatches: 8  
Query Match: 48.83% Indels: 3  
DB: 10 Gaps: 0

US-09-744-313A-1 (1-465) x BG256631 (1-1105)  
QY 217 ProAspValAsnLeuGlyLysIleLeuValProGlyLysLeuMetLysGluLys 236  
Db 1 CCAGATGTAATCTTGGGAAATTTAAATCTGTTCTCGAAACTAATGAAAGAAA 60  
QY 237 GlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLysPro 256  
Db 61 GGTTCAGCATTTGGAACTTTTATCATGAATTTTCAATTAATCTTGTGAGTCTCCAAAGCCT 120  
QY 257 LysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnLysLysLeu 276  
Db 121 AAACCAAGTAGACAGACAGCACTTCTCAGCCCTACTTTCAGAAACCAACAGAAAGCTT 180  
QY 277 PheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsn 296  
Db 181 TTCAATGATCTGTTTAAAAATAATCAAAACCCGCTGCTGAAATAACAGAGAAAGCAAAAT 240  
QY 297 GlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrVal 316  
Db 241 CAGAAATTTATTTATGGAGGTGATGCTAGAGAGAGTCTATGATTACTCTGATGATGTA 300  
QY 317 GlyArgValValPheGlnValProAspTyrPheHisLeuLeuMetGlyThrArgIle 336  
Db 301 GGACGGGTAGTTCAGGGTTCCTGACTGCTTTCATCATCTCTTAATGGCACTCGAATC 360  
QY 337 LeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGln 356  
Db 361 CTCTTTAAAAACACCCCTGGAAATGTATCTACTTCTCAGTGTAACTAGAACAG 420  
QY 357 LeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCys 376  
Db 421 CTATTTTCAGAGACACCCGTTTGGTCTCACTCATTAACCTTCTCAGATGCTATATCTGT 480  
QY 377 GluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGlu 396  
Db 481 GAAACACTGAACTGCTCTCTCCAGATAAGCAAAAGAGCAAAACAGACTTTTGA 540  
QY 397 GluMetMetAsnTyrIleProAspLeuValLysCysIleGlyGluGluThrLysTyr 416  
Db 541 GAAATGATGAATTACATTCAGATCTGTAGTCAAGTGTATTGTTGTAAGAAACCAAGTAT 600  
QY 417 GluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThr 436  
Db 601 GAAACGATCAGACTTCTGTT-GATGGCTTACAGCAACAGTCTCTCAACAGCAGCTGACT 659  
QY 437 TyrValLeuLeuAspIleValIleGlnGluLeuPheProGluLeuAsnLysValGln-L 456  
Db 560 TATGTTTTATCGGACTTGTGATCCGGGAACCCGTTCCAGAGATTCAATAGGTGACACA 719  
QY 456 ysGluValThrSerValThrSerTyr 464  
Db 720 CGGAAGTTACTCTCTGTGACATCTGG 745

RESULT 13  
BG293966 1093 bp mRNA linear EST 21-FEB-2001  
LOCUS 602391002F1 NIH\_MGC\_94 Mus musculus cDNA clone IMAGE:4502737 5',  
DEFINITION mRNA sequence.  
ACCESSION BG293966  
VERSION BG293966.1 GI:13054129  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1093)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory

cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM10372 row: d column: 02  
 High quality sequence stop: 738.  
 Location/Qualifiers

# FEATURES

```

source
1..1093
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:10090"
/clone="IMAGE:4502737"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 94"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI; Site 2: SalI; Cloned unidirectionally; oligo-dT primed. Average insert size 3.3 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH MGC library."
BASE COUNT 353 a 223 c 280 g 237 t
ORIGIN
Alignment Scores:
Pred. No.: 4,19e-117 Length: 1093
Score: 1172.00 Matches: 246
Percent Similarity: 86.15% Conservative: 9
Best Local Similarity: 83.11% Mismatches: 34
Query Match: 48.09% Indels: 8
DB: 10 Gaps: 3

US-09-744-313A-1 (1-465) x BG293966 (1-1093)
QY 19 SerPheGlyLeuSerArgIleGlySerLysIleLysGlyValpPheLysSerThrMet 38
Db 3 TCCTTTGGATCTGACAGATAGGTAGCAAAATTAAGCGCTATTCAAGGTACGCAATG 62
QY 39 GluGlyAlaMetLeuProAsnTyrglyValAlaGluGlyGluAspPheIleGlu 58
Db 63 GAGGGAGCTGTGCTCTTAATACGGGTGGCTGAAGGCGAAGATGACTTTATTAGAA 122
QY 59 GlyIleValValMetGluAspSerProValGluAlaValSerThrProAsnThrPro 78
Db 123 GCGATTGTGTAATGAGGATGACTCTCCAGTAGAGCTGTGAGCACACCTAACACTCT 182
QY 79 ArgAsnLeuAlaAlaTrpLysIleSerIleProTyrglyValAspPheGluAspProSer 98
Db 183 CGAAACCTTGCTCGCTGGAAAATTAGCATTCGTATGTAGACTTTTGAAGATCCCTCC 242
QY 99 SerGluArgLysGluLysLysGluArgIleProValPheCysIleAspValGluArgAsn 118
Db 243 TCTGAAGGAAAGAGAGAGAGAGATTCCTGTGTTTGTATCGATGTTGAAGAAC 302
QY 119 AspArgAlaValGlyHisGluProGluHisTrpSerValTyrglyArgTyrglyLeu 138
Db 303 GACAGACGACGAGTTGGACACGAGCTGAGACTGTGCTGTATAGAGATATCTTGA 362
QY 139 PheTyrglyLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeu 158
Db 363 TTCTATGTTCTTGAATCAAACTTAACAGAAATTCATGGCACAATTTCTGATGCTGCT 422
QY 159 ProSerLysArgIleIleGlyProLysAsnTyrglyPheLeuLysSerLysArgGlu 178
Db 423 CCATCCAAAGGATCATTTGGCCCCCAAAACTATGAGTTTGAAGTCGAAGAGAGAG 482
QY 179 PheGlnGlyTyrglyLeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeu 198
Db 483 TTCAGAGGATCTTGCAGAAACTTGTGCAGACCCCGAGCTGATTAACAGCAGCTTCT 542
QY 199 AlaAspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAsp 218

```

```

Db 543 GCTGACTTCTCTCCCAACGGTGGGAGACACAGTTTCTAGATAAGATATCTCCAGAT 502
QY 219 ValAsnLeuGlyLys-IleIle-LysSerValProGlyLysLeu-MetLysGluLys-Gl 237
Db 603 GTAAATCTTGGAAACATTATAACAGTCTGTCTCTGGGAACTAATTGAAAGAGACGG 562
QY 237 YGlnHisLeuGluProPheIleMetAsnPhelleAsnSerCysGluSerProLysProly 257
Db 663 TCAGCATTCGGAGCCATTTCATCATGAGTTTCTTATTTCTTGGATCTCCAAATCAA 722
QY 257 sProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeu 277
Db 723 CGAGTCTG---GCAGAACTGGGGAAATCTCAGGCCAATTCGAA---ACCAAGAGCTTT 776
QY 277 eAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGl 297
Db 777 CCAGACCGGCTTAGCAATAA-CGGAACCGGGCGGAGAAACCGAGGGGGAACCAATGA 835
QY 297 nAsnTyrglyMetGluValMetThrValGluGlyValTyr 310
Db 836 A---ATACTCAGGAGAGGAAACGGGGACGGGGCGCTAC 872

RESULT 14
BU273460 754 bp mRNA linear EST 26-NOV-2002
LOCUS 603375273F2 CSQCHN53 Gallus gallus cDNA clone ChEST288a20 5', mRNA
DEFINITION
sequence.
ACCSSION BU273460
VERSION BU273460.1 GI:25544410
KEYWORDS EST.
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 754)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22335534
12445392
COMMENT
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology (UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.
Location/Qualifiers
1..754
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="ChEST288a20"
/sex="Female"
/tissue_type="cerebrum"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="CSQCHN53"
/note="Organ: brain; Vector: pBluescript II KS(+); Site_1:
EcoRI; Site_2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with EcoRI
, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS

```





```

Db      302 TTCCAAGAAATATCTACAGAACTTCTGCAGCATCCAGAACTGAGTAATAGTCAACTCTG 361
QY      199 AlaAspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAsp 218
Db      362 GCAGACTTTCTTTCCCTTAATGGTGGGAACACAAATTTCTTGATAGATATCTACCGAT 421
QY      219 ValAsnLeuGlyLysIleLeuSerValProGlyLysLeuMetLysGluLysGlyGln 238
Db      422 GTAATCTTGGGAATAATTATAAACTGTCTCTGAAACTTAATGAAGAGAAAGGTCTAG 481
QY      239 HisLeuGluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysPro 258
Db      482 CATTTGGAACCTTTTATCATGAATTTCAATTTCTGTGAGTCTCCAAAGCCTAAGACA 541
QY      259 SerArgProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsn 278
Db      542 AGTAGACCAAGAACTGACCAATTTCTCAG-CCTACTTCAGAAACACACAGAAGCTTTTCAAT 600
QY      279 AspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluAsnGlyGlnAsnGlnAsn 298
Db      601 GATCTGTTTACAAATAATCAACCCGTGCTGA-AATACAGAGAGAAAGCAACATCAGAAAT 659
QY      299 TyrPheMet-GluValMet---ThrValGluGly-ValTyrAspTyrLeu-MetTyrVal 316
Db      660 TAATTTATCGAGGTGGATTGACTGCTAGAGAGAGGTCTATGATTCATGATGATGATGTA 719
QY      317 GlyArgVal---ValPheGlnValProAspTyrPheHisHisLeuLeuMetGlyThrArg 335
Db      720 AGGACGGGTAAATTTACCGAGTTCCTGAATGGTTAAA-AATTTCTTAATGGGAAT-CGA 777
QY      336 IleLeuPheLysAsnThr 341
Db      778 AATCTCTTTAAAAACACC 795

```

Search completed: January 31, 2004, 05:10:00  
 Job time : 2958 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

DOM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 02:43:12 ; Search time 5257 Seconds  
(without alignments)  
3618.599 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 2437

Sequence: 1 MYLHFCILIFRNTQKRGSEF.....ELFPFLNKVKQKVTSTWSM 465

Scoring table: BLOSUM62  
Xgapop 10.0, Ygapext 0.5  
Fgapop 6.0, Dgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-Q=/cgn2.1/USPTO.spool/US09744313/runat.30012004.113535.4184/app.query.fasta\_1.647  
-DS=GenEmbl -QFMT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313 @CGN 1.1 3508 @runat.30012004.113535.4184 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.ov.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.man.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2437	100.0	1992	6	AX054819 Sequence
2	2387	97.9	1782	10	BC043328 Mus muscu
3	2382	97.7	3145	6	AX512835 Sequence
4	2382	97.7	3145	9	AK000362 Homo sapi
5	2380	97.7	2661	9	AY044865 Homo sapi
6	2380	97.7	3038	9	BC005110 Homo sapi
7	2371	97.3	2925	9	BC046520 Homo sapi
8	2358.5	96.8	3576	9	AK026479 Homo sapi
9	2354	96.6	3616	9	AK095380 Homo sapi
10	1964	80.6	1593	9	AF121863 Homo sapi
11	847	34.8	968	9	AF120561 Homo sapi
12	743	30.5	425	6	BD113855 EST and e
13	634	26.0	358	6	BD026500 Sequence
14	581	23.8	278898	2	AC116713 Mus muscu
15	578	23.7	129010	9	AL589666 Human DNA
16	447	18.3	224551	2	AC111832 Rattus no
17	447	18.3	278375	2	AC130093 Rattus no
18	442.5	18.2	183972	2	EX537259 Dancio rer
19	314.5	12.9	84705	3	CEY48E1B Caenorabd
20	262	10.8	1617	9	AY044656 Homo sapi
21	262	10.8	2720	9	AK095661 Homo sapi
22	249	10.2	454	11	G30543 human STS S
23	229	9.4	1903	9	AK026814 Homo sapi
24	200.5	8.2	1964	9	BC029868 Homo sapi
25	198.5	8.1	2979	9	AF395843 Homo sapi
26	198.5	8.1	6049	9	D87443 Human mRNA
27	192.5	7.9	3040	10	BC004635 Mus muscu
28	192.5	7.9	5760	10	AK122231 Mus muscu
29	188	7.7	2955	6	BD160169 Primer fo
30	188	7.7	2955	9	AK023740 Homo sapi
31	188	7.7	3474	9	AF420470 Homo sapi
32	188	7.7	4129	9	BC045667 Homo sapi
33	184	7.6	1892	9	AK026643 Homo sapi
34	184	7.6	3832	9	BC041844 Homo sapi
35	182.5	7.5	4036	3	AY089224 Drosophil
36	182	7.5	1701	9	AF121862 Homo sapi
37	181	7.4	1472	3	AK117000 Ciona int
38	173	7.1	4879	10	AK122352 Mus muscu
39	168	6.9	34567	8	AF109957 S.pombe c
40	163.5	6.7	1803	10	AF312007 Mus muscu
41	156	6.4	1699	9	AK026637 Homo sapi
42	156	6.4	2065	9	BC005022 Homo sapi
43	156	6.4	2271	9	AF121859 Homo sapi
44	156	6.4	2310	9	AF131214 Homo sapi
45	156	6.4	4175	9	AF172847 Homo sapi

# ALIGNMENTS

RESULT 1

201	Qy	PhelLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsn	220
716	Db	TTTCTTTCCCTTAATGGTGGGAACAACAATTTCTTGATAGACTACTACCGATGTAAAT	775
221	Qy	LeuGlyLysIleLeuLysSerValProGlyLysLeuMetLysGluLysGlyGlnHisLeu	240
776	Db	CTTGGGAAATATAAAATCTGCTCTGGAAAACCTAATGAAAGAGAAAGGTACGATTTG	835
241	Qy	GluProPheIleMetAsnPheIleAsnSerCysGluSerProLysProLysProSerArg	260
836	Db	GAACCTTTATCATGAATTTCAATTAATTTCTTGAGTCTCCAAAGCCCTAAACCAAGTAGA	895
261	Qy	ProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu	280
896	Db	CCAGACTGACCAATCTCGACCTTACTCAGAAAACACACAGAGAGCTTTCAATGATCTG	955
281	Qy	PhelysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPhe	300
956	Db	TTTAAAAATAATCAACACCGTCTGAAAAATACAGAGAGAAAGCAAAATCAGAAATTAATTTT	1015
301	Qy	MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal	320
1016	Db	ATGAGGTGATGACTCTAGAGAGAGTCTATGATTACCTGATGTATAGGACGGGTAGTT	1075
321	Qy	PheGlnValProAspTrpLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsn	340
1076	Db	TTCCAGGTTCCTGACTGGGCTTCATCATCTCTTAATGGGAACTCGAATCCTCTTTAAAAAC	1135
341	Qy	ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGlu	360
1136	Db	ACCTCGAAATGTAATCTGATCTTACTATCTTCACTGTAACCTAGAACAGCTATTTTCAGGAG	1195
361	Qy	HisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGlu	380
1196	Db	CACCGTTTGGTCTCACTCATCAACACTTCTCAGAGATGCTATATTCTGTGAAAAACACTGAA	1255
381	Qy	ProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsn	400
1256	Db	CCTCGCTCTCTCCAAAGATAGCAAAAGAGCAAAACACAGACTTTTGAAGAAATGATGAAT	1315
401	Qy	TyrIleProAspLeuValLysCysIleGlyGluGluThrLysTyrGluSerIleArg	420
1316	Db	TACATTCAGATCTGTTAGTCAAGTGTATTGGTGAAGAAACCAAGTATGAAGCATCAGA	1375
421	Qy	LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu	440
1376	Db	CTTCTGTTGATGGCTTACAGCAACAGTACTCAACAGCAGCTGACATTATGTTTATTG	1435
441	Qy	AspIleValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGluValThrSer	460
1436	Db	GACATCTGTGATACAGGAACCTGTTTCCAGAGCTCAATAGGTACAAAGAGAAAGTTACCTCT	1495
461	Qy	ValThrSerTrpMet	465
1496	Db	GTGACATCTTGGATG	1510
RESULT 2			
BC043328			
LOCUS			
DEFINITION			
IMAGE:5400157), complete cds.			
ACCESSION			
BC043328			
VERSION			
BC043328.1			
KEYWORDS			
SOURCE			
ORGANISM			
Mus musculus (house mouse)			
Mus musculus			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
1 (bases 1 to 1782)			
Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,			
Klauser,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,			
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,			
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,			

ERIPVFCIDVERNDRAVGHPEHVSRRYVLEFVLBSKLTETFGTTPDQALPSKRI  
IGPKNYEFLKSRREFQYLOKLVQHPLESLQADFLSPNGSETQFLDKLPDNL  
GKIISVPGKMKRQGLHLEFIMFINSCEPKSPKSPRLTILSP7SENNKLPND  
LFPKNNARENTKONONYFMEVTVGVYDYLWVGVRVQVQVFWLHLLMGRIL  
FKNTLEMYDYLOKLCQLQOEHLVSLITLRLDAI CENTEPESLODKOKGAKTF  
REMMYIPDLVKIGERTKESIRLLPDGLQOQVYNKQLTIVLLDIVIQELFPELNK  
VQKEATSMTSMW

BASE COUNT 567 a 339 c 380 g 496 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1,1e-198 Length: 1782  
Score: 2387.00 Matches: 453  
Percent Similarity: 99.14% Conservative: 8  
Best Local Similarity: 97.42% Mismatches: 4  
Query Match: 97.95% Indels: 0  
DB: 10 Gaps: 0

US-09-744-313A-1 (1-465) x BC043328 (1-1782)

Qy 1 MetTyrlleuHleHisPheCysLeuLlePheArgAsnThrGlnLysArgGlyGluSerPhe 20  
Db 219 ATGTACTGATGCGTATTGTTTAAATTTAGAGCAGCACAAAAGAGAGAGAGATGTTT 278  
Qy 21 GlyLseArgIleGlySerLysLleLysGlyValPheLysSerThrThrMetGluGly 40  
Db 279 GGAATCAGCAGATAGTAGCAAAATTAAGGCGTATTCAAGAGTAGCAGCAATGAGGGA 338  
Qy 41 AlaMetLeuProSerTyrlGlyValAlaGluGlyGluAspPheLleGluGlyLle 60  
Db 339 GCTGTGCTGCTTAATTACGGGGTGGCTGAAGCGGAGATGACTTTATTGAAGAAGGATT 398  
Qy 61 ValValMetGluAspSerProValGluAlaValSerThrProAsnThrProArgAsn 80  
Db 399 GTGTAATGAGGATGACTCTCCAGTAGAGCTGTGAGCAGACACTAACCTCTCCGAAC 458  
Qy 81 LeuAlaAlaTrpLysLleSerIleProTyrlValAspPheGluAspProSerSerGlu 100  
Db 459 CTTGCTGCTGGAAAATTAGCATTGCTGATGAGACTTTTTCAGAGATCTCTCTCGAA 518  
Qy 101 ArgLysGluLysLysGluArgLleProValPheCysIleAspValGluArgAsnAspArg 120  
Db 519 AGGAAGAAGAAGAAGAGAGAGATTCCTGCTGTTTGTATCGATGTTGAAGAAGACAGACA 578  
Qy 121 ArgAlaValGlyHisGluProGluHisTrpSerValTyrlArgArgTyrlLeuGluPheTyrl 140  
Db 579 CGAGCAGTTGCACACGAGCTGAGCATTGCTGCTCTATAGAAGATATCTTGAATCTAT 638  
Qy 141 ValLeuGluSerLysLeuThrGluPheHisGlyAlaPheProAspAlaGlnLeuProSer 160  
Db 639 GTTCTTGATCAAACTAACAGAAATTTTCATGGGCACATTTCTCTGATGCTCAGTCTCCATCC 698  
Qy 161 LysArgLleLleGlyProLysAsnTyrlGluPheLysSerLysArgGlyGluGluPheGln 180  
Db 699 AAAAGGATCATTTGCCCCAAAATACTAGTTTGTAGTCGAGAGAGAGAGATTCAG 758  
Qy 181 GluTyrlLeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAlaAsp 200  
Db 759 GAGTATCTGCAGAACTTGTGCACACCCGAGCTGAGTAACAGCAGCTTCTGGCTGAC 818-  
Qy 201 PheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysLleLeuProAspValAsn 220  
Db 819 TTTCTCTCCCAACGGTGGGAGACACAGTTTCTAGATAAGATATCTCAGATGTAAT 878  
Qy 221 LeuGlyLysLleLleLysSerValProGlyLysLeuMetLysGlyLysGlyGlnHisLeu 240  
Db 879 CTGGAAAATAATAAGTCTGTTCTCCGGGAACTAATGAAGAGAAAGGTCAGCATTTG 938  
Qy 241 GluProPheLleMetAsnPheLleAsnSerCysGluSerProLysProLysProSerArg 260  
Db 939 GAGCATTTCATCAGATTTCATTAATTTCTTGGCAATCTCCCAACCTAACCGAGTCGG 998  
Qy 261 ProGluLeuThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeu 280

Diatchenko, L., Marusina, K., Farmer, A.A., Rubin, G.M., Hong, L.,  
Stapleton, M., Soares, M.B., Bonaldo, M.F., Casavant, T.L.,  
Carrinci, P., Brownstein, M.J., Usdin, T.B., Toshiyuki, S.,  
Santini, P., Prange, C., Raha, S., Loquellano, N.A., Peters, G.J.,  
Abramson, R.D., Mullahy, S.J., Bosak, S.A., McEwan, P.J.,  
McKernan, K.J., Male, J.A., Gunaratne, P.H., Richards, S.,  
Worley, K.C., Hale, S., Garcia, A.M., Gay, L.J., Hulyk, S.W.,  
Villalon, D.K., Muzny, D.M., Sodergren, E.J., Lu, X., Gibbs, R.A.,  
Fahey, J., Helton, E., Kettman, M., Madan, A., Rodriguez, S.,  
Sanchez, A., Whiting, M., Madan, A., Young, A.C., Shevchenko, Y.,  
Bouffard, G.G., Blakesley, R.W., Touchman, J.W., Green, E.D.,  
Dickson, M.C., Rodriguez, A.C., Grimwood, J., Schmutz, J., Myers, R.M.,  
Butterfield, Y.S., Krzyzinski, M.I., Skalska, U., Smalilus, D.E.,  
Schmerch, A., Schein, J.E., Jones, S.J. and Marra, M.A.  
Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences  
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)

22388257  
12477932  
2 (bases 1 to 1782)  
Straussberg, R.  
Direct Submission  
Submitted (09-JAN-2003) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)  
Tissue Procurement: The Cepko Laboratory  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@bcgsc.bc.ca](mailto:info@bcgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Boedet, Yaron Butterfield,  
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
Leticia Hsiao, Martin Krzyzinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
Schein, Duane Smalilus, Michael Smith, Lorraine Spence, Jeff Stott,  
Michael Thorne, Miranah Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 86 Row: P Column: 11  
This clone was selected for full length sequencing because it  
passed the following selection criteria: Hexamer frequency ORF  
analysis.

FEATURES  
source Location/Qualifiers  
1..1782  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="MGC:49424 IMAGE:5400157"  
/tissue\_type="Eye, retina, mouse strain C57Bl/6"  
/clone\_lib="NIH\_MGC\_94"  
/lab\_hosts="DH10B"  
/notes="Vector: pCMV-SPORT6"  
1..1782  
/gene="C330035N22R1k"  
/notes="synonym: B830022K16"  
/db\_xref="LOCUSID:244962"  
/db\_xref="MGI:2443156"  
228..1616  
/codon\_start=1  
/product="C330035N22R1k protein"  
/protein\_id="AAH43328.1"  
/db\_xref="GI:27694049"  
/db\_xref="LocusID:244962"  
/translations="MRICLFRSTQKRGSGFISRGISKGVFKSTTWGAVLPNYG  
VAEGEDDFIEGIVMEDDSPVEAVSTPNTPRNLAANKISIPVDFPDPSSERKEKK

TITLE  
JOURNAL  
MEDLINE  
PUBMED  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REMARK  
COMMENT

999 CCAGAACTGACGATTTCTAGCCCACTTCCGAAAACACAGAAAGCTTTTCAACGATCTG 1058  
281 PheLysAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnThrPhe 300  
1059 TTTAAGAAATATGCAAAACCGCTGAGACACAGAGCGAAGCAAAATCAGAACTACTTC 1118  
301 MetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValVal 320  
1119 ATGAGGTGATGCTGTGAGCGAGTCTATGATACCTGATGATGATGATGAGCGAGTAGTT 1178  
321 PheGlnValProAspThrLeuHisLeuLeuMetGlyThrArgLysLeuPheLysAsn 340  
1179 TTTCAAGTCCAGACTGGCTTCATCTCTTAATGGGAACTCGAATCTCTTTAAGAAC 1238  
341 ThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLeuLeuGluGlnLeuPheGlnGlu 360  
1239 ACCCTGGAAATGACACAGACTACTCTCCAGTGCAGCTGAGAGAGCTGTTTCAAGNG 1298  
361 HisArgLeuValSerLeuLeuThrLeuLeuArgAspAlaLeuPheCysGluAsnThrGlu 380  
1299 CACCGGTGCTCTCCCTCATCACCCTCTCAGAGATGCTATATTTTGTGAAAATACCTGAA 1358  
381 ProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsn 400  
1359 CCTCGCTCTCTCCAGATAAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAAT 1418  
401 TyrIleProAspLeuLeuValLysCysLysGlyGluThrLysTyrGluSerIleArg 420  
1419 TACATTCAGATCTGATAGTCAAGTGTATCGGTGAAGAACCAAGTACCAAGACATCAGA 1478  
421 LeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeu 440  
1479 CTCCTGTTGATGCTGAGCAGCAGCAGCTGCTCAACAGAGCTGACCTATGTTGTCG 1538  
441 AspIleValIleGlnLeuPheProGluLeuAsnLysValGlnLysGluValThrSer 460  
1539 GACATTGTGATACAGGAGCTGTTTCCAGAGCTAAATAAGGTACAAAAGGAGCTACCTCC 1598  
461 ValThrSerThrMet 465  
1599 ATGACATCCTGGATG 1613

RESULT 3  
AX512835 3145 bp DNA linear PAT 03-OCT-2002  
LOCUS Sequence 4 from Patent WO02062839.  
DEFINITION AX512835  
ACCESSION AX512835  
VERSION AX512835.1 GI:23504019  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Daenen, M.J., Cleutjens, C.B. and Zaman, G.J.  
TITLES Markers of unstable atherosclerotic plaques  
JOURNAL Patent: WO 02062839-A 4 15-AUG-2002;  
Universiteit Maastricht (NL)  
FEATURES  
source  
1. 3145  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 985 a 570 c 659 g 931 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6,03e-198 Length: 3145  
Score: 2382.00 Matches: 454  
Percent Similarity: 100.00% Conservative: 2  
Best Local Similarity: 99.56% Mismatches: 0  
Query Match: 97.74% Indels: 0

DB: 6 Gaps: 0  
US-09-744-313A-1 (1-465) x AX512835 (1-3145)  
Qy 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29  
Db 1598 TTTCCGGAACACACAGAAAAGGGGAGAAATCATTTGGAAATCAGCAGAAATAGTAGCAAAAT 1657  
Qy 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
Db 1658 AAAGGAGTATTTCAGAAAGTACCAACATGGAGGGAGCTATGTTGCCTTAATATATGTTAGCT 1717  
Qy 50 GluLysGluAspPheIleGluGlyValValMetGluAspSerProVal 69  
Db 1718 GAAGGTGAAGATGATTTTATTAAGAGAGTATTTGTTAATGGAAGATGATTTCTCCAGTG 1777  
Qy 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro 89  
Db 1778 GAGGCTGTGAGCACACCTAATATCTCCCGAAACCTTGCTGCTAGTGAATAATTAGCATTTCCA 1837  
Qy 90 TyrValAspPhePheGluAspProSerSerGluArgLysGlnLysLysGluArgIlePro 109  
Db 1838 TATGTAGACTTTTGTAGGATCCCTCTCTGAAAGGAGGAGAAAAGAAAGAAATTCCT 1897  
Qy 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
Db 1898 GTGTTTGTATGTATGTGAAGAAATGATAGAGAGCAGTTGGACAGAGCTGACAT 1957  
Qy 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
Db 1958 TGGTCTGTATAGAGATATCTGAATTTCTATGACTTGAATCAAAATCAACAGAAATTT 2017  
Qy 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
Db 2018 CATGTGCTATTTCTGATGCCAGCTTCTTCTAAGAGAGTCAATTTGGCCCCCAAAATAT 2077  
Qy 170 GluPheLeuLysSerLysArgGluGluPheGlnGlyTyrLeuGlnLysLeuGlnHis 189  
Db 2078 GAATTTCTTAAAGTCAAGAGAGGAGTTCAGAAATATCTACAGAACTTCTGACGAT 2137  
Qy 190 ProGluLeuSerAsnSerGlnLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
Db 2138 CCGAAGCTGAGTATATAGTCACTTCTGGCAGACTTTCTTTCCCTTAATGGTGGGGAACA 2197  
Qy 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValPro 229  
Db 2198 CAATTTCTTGATAGATAGTACTACAGATGTAATCTTGGGAAAATATATAAAATCTGTTCCT 2257  
Qy 230 GlyLysLeuMetLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
Db 2258 GGAAGAACTAATGAAAGAGAGAGTCCAGCATTTGGAACCTTTTATCATGAATTTTCAAT 2317  
Qy 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
Db 2318 TCTTTGTAGTCTCCAAAGCCTTAACCAAGTAGACAGAACTGACCATTTCTCAGCCCTACT 2377  
Qy 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
Db 2378 TCAGAAAACACACAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTCTGAA 2437  
Qy 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
Db 2438 AATACAGAGAGAGAAAGCAAAATCAGAAATATTTTATGGAGGTGATGACTGTAGAAGGATC 2497  
Qy 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspThrLeuHis 329  
Db 2498 TATGATTACCTGATGTATGTAGGCGGTAGTCTTCCAGATTTCTCCAGATTCCTGATTCAT 2557  
Qy 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
Db 2558 CTCCTTAATGGGAACCTCGAATCTCTTTTAAACACCTCGGAAATGATATGATGATGAT 2617  
Qy 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeu 369

Db 2618 CTTGAGTGAACATGAGACAGCTATTTTCAGAGACACGGTTGGTCTCCTCATTAACACTT 2677  
Qy 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
Db 2678 CTCAGAGATGCTATATCTTGTGAAACACATGAACCTCGCTCTCTCCAAAGTAAGCAAAA 2737  
Qy 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
Db 2738 GGAGCAAAACAGACTTTTGAAGAAATGATGAATACATTCAGATCTGTAGTCAAGTGT 2797  
Qy 410 IleGlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
Db 2798 ATTGGTGGAGAAACCAAGATATGAAGCATCAGACTTCTGTTTGAATGGCTTACAGCAACCA 2857  
Qy 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
Db 2858 GTACTCAACAGCAGCTGACTTAATGTTTATGGACATTTGGATACAGGACTGTTTCCA 2917  
Qy 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
Db 2918 GAGCTCAATAGGTACAAAGAGAGTACCTCTGTGACATCTGGATG 2965

AK000362 3145 bp mRNA linear PRI 22-FEB-2000  
LOCUS Homo sapiens cDNA FLJ20355 fis, clone HEP15804, highly similar to  
DEFINITION AF121863 Homo sapiens sorting nexin 14.  
ACCESSION AK000362  
VERSION AK000362.1 GI:7020397  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (sites)  
AUTHORS Kawakami, T., Noguchi, S., Itoh, T., Shigeta, K., Senba, T.,  
Matsumura, K., Nakajima, Y., Mizuno, T., Morinaga, M., Ota, T.,  
Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,  
Nakamura, Y., Isogai, T. and Sugano, S.  
NEDO human cDNA sequencing project  
Unpublished  
TITLE NEDO human cDNA sequencing project  
JOURNAL 2 (bases 1 to 3145)  
AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,  
Shibahara, T., Tanaka, T. and Nakamura, Y.  
Direct Submission  
JOURNAL Submitted (15-FEB-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,  
Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,  
Tel:81-3-5449-5286, Fax:81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing: Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing: Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

FEATURES Location/Qualifiers  
1..3145  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="HEP15804"  
/cell\_line="HepG2"  
/cell\_type="hepatoma"  
/clone\_lib="HEP"  
/notes="cloning vector pME18SFL3"  
misc\_feature 1..3145  
/note="highly similar to AF121863 Homo sapiens sorting  
nexin 14"  
BASE COUNT 985 a 570 c 659 g 931 t  
ORIGIN

Alignment Scores: 6.03e-198 Length: 3145  
Pred. No.: 2382.00 Matches: 454  
Score: 2382.00 Conservativeness: 2  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 99.58% Indels: 0  
Query Match: 97.74% Gaps: 0  
DB: 9  
US-09-744-313a-1 (1-465) x AK000362 (1-3145)  
Qy 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIle 29  
Db 1598 TTTTCGAAACACACAGAAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGGTAGCAAAAT 1657  
Qy 30 LysGlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAla 49  
Db 1658 AAAGGAGTATTGAGAGTACCACAAATGGAGGAGGCTATGTTGCCCTAAATATGGTGTAGCT 1717  
Qy 50 GluGlyGluAspAspPheIleGluGluGlyIleValMetGluAspAspSerProVal 69  
Db 1718 GAAAGGTGAAGATGATTTTATTCAGAGAGGTATTTGTTATTCAGAGATGATTTCTCCAGTG 1777  
Qy 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIlePro 89  
Db 1778 GAGGCTGTGAGCACACCTAATACTCCCGAAACCTTGTCTGATGGAATAATAGCATTTCCA 1837  
Qy 90 TyrValAspPhePheGluAspProSerSerGluArgLysGluLysGluLysGluLysGlu 109  
Db 1838 TATGTAGACTTTTTCAGGATCCCTCTCTGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1897  
Qy 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129  
Db 1898 GTGTTTGTATTTGATGTTGAAGAAATGATAGAGAGAGGAGTGGACACAGGCTGACAT 1957  
Qy 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
Db 1958 TGGTCTGTCTATAGAGAGATATCTTGAATCTATGTAATCAAACTAAACAGAAATTT 2017  
Qy 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
Db 2018 CATGGTGCAATTTCTCTGATGCCAGCTTCTCTTCTTAAGAGGATCATTTGGCCCCCAAAATAT 2077  
Qy 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHis 189  
Db 2078 GAATTTCTTAAGTCAAGAGGAGAGAGTTCACAGAAATATCTACAGAAATCTTCGAGCAT 2137  
Qy 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
Db 2138 CCAGAACTGAGTAAATAGTCAACTTCTGGCAGACTTTCTTCCCTTAATGGTGGGAAACA 2197  
Qy 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValPro 229  
Db 2198 CAATTTCTTGAAGATACCTACAGATGTAATCTTTGGGAAATATATAAATCTGTTCCT 2257  
Qy 230 GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheLeuAsn 249  
Db 2258 GGAATACTAATGAAGAGAAAGGTGAGCATTTGGAACCTTTTATCATGAATTTCAATTAAT 2317  
Qy 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
Db 2318 TCTTTGTGAGTCTCCAAAGCCTTAACCAAGTAGACCAAGTACCATCTTCAGCCCTACT 2377  
Qy 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
Db 2378 TCAGAAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGGTCTGAA 2437  
Qy 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309  
Db 2438 AATACAGAGAGAAAGCAAAATCAGAAATATTTTATGAGGAGGTGATGATCTGTAGAGGAGTC 2497  
Qy 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrLeuHis 329  
Db 2498 TATGATTACCTGATGTATGTAGGACGGGTAGTTTTCAGATTTCTGACTGCTTCATCAT 2557



BASE COUNT 855 a 490 c 534 g 782 t

ORIGIN

Alignment Scores:

Pred. No.: 7,34e-198 Length: 2661

Score: 2380.00 Matches: 455

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 97.66% Indels: 0

DB: 9 Gaps: 0

US-09-744-313a-1 (1-465) x AY044865 (1-2661)

QY 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyLeuSerArgIleGlySerLysIleLys 30

DB 1294 AGGAACACACAGAAAAGGGAGGAATCATTTGGAATCAGCAGAAATAGGTAGCAAAATATAA 1353

QY 31 GlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGlu 50

DB 1354 GGAGTATTCAAAAGTACCACAAATGGAGGAGCTATGTCCTTAATATGCTAGTGTGAA 1413

QY 51 GlyValAspPheIleGluGlyValValMetGluAspAspSerProValGlu 70

DB 1414 GGTGAAGATGATTTTATGAAGAGGATGTTGTTGTAATGGAAGATGATTTCCAGTGGAG 1473

QY 71 AlaValSerThrProAsnThrProArgAsnLeuAlaAlaTrpLysIleSerIleProTyr 90

DB 1474 GCTGTGAGCACACTAATCTCCCGAAACCTTGCTGCATGGAAATATAGCATTCATAT 1533

QY 91 ValAspPheGluAspProSerSerGluArgLysGluLysLysLysLysLysLysLys 110

DB 1534 GTAGACTTTTTCAGGATCCCTCTCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1593

QY 111 PheCysIleAspValGluArgAsnAspArgAlaValGlyHISgluProGluHisTyr 130

DB 1594 TTTTGTATTCATGTTGAAGAAATGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1653

QY 131 SerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150

DB 1654 TCTGTCTATAGAGATATCTTGAATCTATGATCTTGAATCAAACTAACAGAAATTCAT 1713

QY 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu 170

DB 1714 GGTGCAATTCCTGATGCCAGCTTCCTCTAGAGAGATCATTCGCCCAAAATATGAA 1773

QY 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190

DB 1774 TTTCTAAAGTCAAG 1833

QY 191 GluLeuSerAsnSerGlnLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210

DB 1834 GAATGAGTAAATAGTCAACTTCCTGAGAGATTTTCTTCCCTTAATGAGTGGGAGAGAG 1893

QY 211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValProGly 230

DB 1894 TTTCTTGTATAGATACATCCAGATGTAATCTTGGGAAATATATAAATCTGTTCTCTGA 1953

QY 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheLeuAsnSer 250

DB 1954 AAACCTAAAG 2013

QY 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270

DB 2014 TGTGAGTCTCAAG 2073

QY 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290

QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349

DB 2558 CTTCTAAATGGGAACTCGAATCTCTTTTAAACACCCCTGGAAATGATATGATATAT 2617

QY 350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeu 369

DB 2618 CTTCTAGTAACTAGAACAGCTATTTTCAGAGACACCGTTTGGTCTCACTATAACACTT 2677

QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389

DB 2678 CTTCTAGATGCTATATCTCTGTGAAACACCTGAACTGCTCTCTCCAGATAGCAAAA 2737

QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409

DB 2738 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTCATCCAGATCTCTTGTAGTCAAGTGT 2797

QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429

DB 2798 ATTGTGAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCAACCA 2857

QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449

DB 2858 GTACTCAACAGAGCTGACTATGTTTATGGACATTTGATACAGAGAACTGTTTCCA 2917

QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465

DB 2918 GAGCTCAATAGGTACAAAGGAAGTTACCTCTGTGACATCTTGATG 2965

RESULT 5

AY044865

LOCUS 2661 bp mRNA linear PRI 09-SEP-2001

DEFINITION Homo sapiens sorting nexin 14 (SNX14) mRNA, complete cds.

ACCESSION AY044865

VERSION AY044865.1 GI:15529063

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2661)

Hong, W.

The complete coding region of SNX14

Unpublished

REFERENCE 2 (bases 1 to 2661)

Hong, W.

Direct Submission

Submitted (11-JUL-2001) Membrane Biology Laboratory, Institute of Mol & Cell Biol, 30 Medical Drive, Singapore 117609

Location/Qualifiers

1..2661

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

1..2661

/gene="SNX14"

/note="contains PX domain"

/codon\_start=1

/product="sorting nexin 14"

/protein\_id="AAK97796.1"

/db\_xref="GI:15529064"

/translation="MQQKQRLDVGREICRQYPLFCFLLCLLSAALLNRYVIHILMFWSVAGVTFYCSLGDPSLLPNIPFTIKYKPKQLGLOELFPGHSCAVCGKVC

KRHPSPILLNRYQWDLKLSSKVDASLSEVDIPSLITKKLKAAMKHIEVIVKARQK

VKNTFEQQALAEYGEPLHVALRSDELHTELLFPYILPPKATDCRSUTLL

IRETLGSVPLSPDLADPTVNHLLIIPIDDSPEPKATEPASPVLPPFLQKFAEPN

KPSVLKLELQREOQDLRFWMFLKQGAHVLFQCLTVEEFNDRILRPELSNDE

MLSHLELQTYKTCIDESIDKTRDPFIVERIORIAEGPYIDVVKLOTMRCLFEAY

EHVLSLENVTFMFWCHSDRYFQLRLGASPTNSKLNENTQKCESGIGRIQSKI

KGVKFTMTSGMAGVAGEDDFIEGIVMHDDSPVEAVSTNTPTNRLNAWKIS

IPYVDFDFFSSRKEKKERIPVCFIDVNRDRAVGHGPEHWSVYRYLLEFVLESK

source

gene

CDS

2074 GAAACACACAGAGCTTTTCAATGATCTGTTTAAATAATTAATGCAACCGTGCTGAAAT 2133

291 ThrGluArgGlyLeuGlnAsnThrPheMetGluValMetThrValGluGlyValTyr 310

2134 ACAG 2193

311 AspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTyrPheHisLeu 330

2194 GATTACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2253

331 LeuMetGlyThrArgIleLeuPheLeuAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350

2254 TTAATGGAACTCGAATCTCTTTTAAACACACCTCGAATGATGATGATGATGATGATG 2313

351 GlnCysLeuLeuGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeu 370

2314 CATGTGAACTGAAACAGCTTTTTCAGGAGCACCGTTTGGTCTCACTCAACACTTCTC 2373

371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspGlyGlnGly 390

2374 AGAGATGCTATATCTTGTGAACACATGAACTCGCTCTCCAGATAAGCAAAAGGA 2433

391 AlaGlyGlnThrPheGluGluMetMetAsnTyrIlePheAspLeuLeuValGlyCysIle 410

2434 GCAAAACAGACTTTTGAAGAAATGATGAATTAATTAATTAATTAATTAATTAATTA 2493

411 GlyGluGluThrTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnProVal 430

2494 GGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTGTGATGCTTACAGCAACAGTA 2553

431 LeuAsnGlyGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGlu 450

2554 CTCACAGAGAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2613

451 LeuAsnGlyValGlnGlyGluValThrSerValThrSerTrpMet 465

2614 CTCATAGGTACAAAGAGAGTACCTCTGTGATGATGATGATGATGATGATGATGATG 2658

RESULT 6

BC005110 3038 bp mRNA linear PRI 12-JUL-2001

LOCUS Homo sapiens, clone MGC:13217 IMAGE:3959086, mRNA, complete cds.

DEFINITION BC005110

ACCESSION BC005110

VERSION BC005110.1 GI:13477272

KEYWORDS MGC.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 3038)

Strausberg, R.

Direct Submission

Submitted (26-MAR-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,

USA

NIH-MGC Project URL: <http://mgc.nci.nih.gov>

Contact: MGC help desk

Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)

Tissue Procurement: ATCC

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Genome Sequence Centre,

BC Cancer Agency, Vancouver, BC, Canada

[info@cgsc.bc.ca](mailto:info@cgsc.bc.ca)

Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,

Suzanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,

Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo

Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven

Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline

Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Scott,

Michael Thorne, Miranada Tsai, Nataşja van den Bosch, Jill Vardy,

George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAL Plate: 18 Row: m Column: 23

This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis.

## FEATURES

source

Location/Qualifiers

1..3038

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clone="MGC:13217 IMAGE:3959086"

/tissue\_type="Placenta, choriocarcinoma"

/clone\_lib="NIH\_MGC\_21"

/lab\_host="DH10B-R"

/note="Vector: pOTB7"

189..2849

/codon\_start=1

/product="Unknown (protein for MGC:13217)"

/protein\_id="AAH05110.1"

/db\_xref="GI:13477273"

/translation="MGOKLQRLRLDVGREICRQYPLFCFLLLCLSAASLLNRYIHI

LMFWSFVAGVTFYCSLGLPDLNFIFFTIKYKPKQLGLOELFPQGHSCAVCGKVK

KRHRPSULLENYQWLDLKSISKVDASLSVDIPSITKLLKAAWKHIEVTYKARQK

VKTFTQQAALBEYEGELHVALRRDELHRLKUTELLFPYILPPLPPTDCLSLIL

IRETLGSLFPLSLDFLDPDTVNHLLIIIFDDSPPEKATEPASPPLVPFLQKFAEPNN

KPSVLKLEUKOIREQDLDFRMNFKQEGAVHLOFCLTVEEFNDRIQLMRCLFRAY

MLSHRELKITYKTCIDESIDKIDPFIIEIQRARGPIYIDVVKLQTMRCILFRAY

KGVKFTSTMSGAMLPTYGVAEGDDFIEEGIVNEDDSFVEASTVPTNPNLAWKIS

IPVYDFEDFSRSKKERKIPVFCIDVENDRRVAGHGEPEHMSVRYHLYEFVYLSK

LTBPHGAPPAQPLSKRIIGPKNYEFLKSRBEFQYLOKLOHPLELSNQLADPLS

PNGESTOFLDKILPDVNLGKIISKVPGKLMEKGHLEPFIIMPFINSCSPKPKPSRP

ELTILSPSTENKKLNDLPKNANRAENTERKONONYFNEVMTVEGVYDLYMYGRV

VFOVPDWHLLMGTRILFNKNTLMDYDITLQCKLEQLFOEHLRLVSLITLLRDAIFCE

NTEPRSLQDKQKAKQTFEEMNYPIDLVKCIQEETKYSIRLLPGLQOQVNLKOL

TYVLLDIVIQELFELNKNVKEVTSVTSNM"

BASE COUNT 949 a 564 c 642 g 883 t

ORIGIN

## Alignment Scores:

Pred. No.: 8,63e-198 Length: 3038

Score: 2380.00 Matches: 455

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 97.66% Indels: 0

DB: 9 Gaps: 0

US-09-744-313A-1 (1-465) x BC005110 (1-3038)

QY 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIleIys 30

Db 1482 AGGAACACACAGAAAGGGAGAAATCATTTGGAAATCAGCAGANTAGGTAGCAAAATATAA 1541

QY 31 GlyValPheLysSerThrThrMetGluGlyValMetLeuProAsnTyrGlyValAlaGlu 50

Db 1542 CGAGTATTCAAAGATACCACCAATGAGGAGACTTGTTCCTTAATTATGTTAGTGTAGTGAA 1601

QY 51 GlyGluAspPheIleGluGluGlyIleValValMetGluAspAspSerProValGlu 70

Db 1602 GGTGAAGATGATTTTATTGAAGAGGATTTGTTGAATGGAAGATGATTTCCAGTGGAG 1661

QY 71 AlaValSerThrProAsnThrProArgAsnLeuAlaIleThrLysIleSerIleProTyr 90

Db 1662 GCTGTGAGCACACCTTAATCTCCCGAAACCTTGTGCAATGCAAAATATAGCATTTCCATAT 1721

QY 91 ValAspPhePheGluAspProSerSerGluArgGlyGluLysGluArgGluArgGluProVal 110

Db 1722 GTAGACTTTTGTAGGATCCCTCTCTCGAAAGGAGGAGAGAAAGAAAGAAATTTCTCTGTG 1781

QY 111 PheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTrp 130

1782 TTTTGTATTGATGTTCAAGAAATGATAGACAGCAGTGGACACGACCTGGAACATTGG 1841  
 QY 131 SerValTyrArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
 Db 1842 TCTGTCTATAGAGATATCTTGAATCTATGACTTGAATCAAACTAACAGATTTCAT 1901  
 QY 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleLeuGlyProLysAsnTyrGlu 170  
 Db 1902 GGTGCATTTCTGTATGCCAGCTTCTTCTAAGAGATCATTTGGCCCAAAATATATGAA 1961  
 QY 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuLeuGlnHisPro 190  
 Db 1962 TTCTTAAATCAAGAGGGAAGATTCCTCAAGAAATATCTACAGAACTTCTTCGAGATCCA 2021  
 QY 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
 Db 2022 GAATGAGTAATAGTCAACTTCTGCGACACTTTTCCCTTAATGTTGGGGAACACAA 2081  
 QY 211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValProGly 230  
 Db 2082 TTCTTGTATAGATATCTACAGATGTAATCTTGGGAATATATAAATCTGTCTCTGGA 2141  
 QY 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleValMetAsnPheIleAsnSer 250  
 Db 2142 AACTAATGAAGAGAAAGGTACAGCTTTGGACCTTTTATCATGAATTCATTAATCT 2201  
 QY 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
 Db 2202 TGTGAGTCTCCAAAGCCCTAAACCAAGTAGACCACTGACCACTTCAGCCCTACTTCA 2261  
 QY 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsn 290  
 Db 2262 GAAACCAACAGAAAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTGTGAAT 2321  
 QY 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
 Db 2322 ACAGAGAGAAAGCAATACAGATTAATTTATGGAGGTGATGACTGTAGAAGGAGTCTAT 2381  
 QY 311 AspTyrLeuMetTyrValGlyArgValPheGlnValProAspTrpLeuHisLeu 330  
 Db 2382 GATTACCTGATGATGATAGGACGGGTAGTTTCCAGGTCTCTGACTGGCTTCATCATCTC 2441  
 QY 331 LeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
 Db 2442 TTAATGGGAACCTCGAATCTCTTTAAACCAACCTCGGAATGATATCTGATTACTATCT 2501  
 QY 351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeu 370  
 Db 2502 CAGTGTAAACCTAGAACAGCTATTTCCAGGACCGCTTTGGTCTCACTCATACACTCTC 2561  
 QY 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
 Db 2562 AGAGATGCTATATTTCTGTGAAAACACTGAACCTGCTCTCTCCAGATAAGCAAAAGGA 2621  
 QY 391 AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
 Db 2622 GCAAAACAGACTTTTGAAGAAATGATGATTAATTCAGATCTGTTAGTCAAGTGATTT 2681  
 QY 411 GlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProVal 430  
 Db 2682 GGTGAAGAACCAAGATGAAAGCATCAGACTTCTGTTGATGGCTTACACCAACAGTA 2741  
 QY 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGlu 450  
 Db 2742 CTCACAGACAGCTGACTTATGTTTATTTGGACATTTGTGATACAGAACTGTTTCCAG 2801  
 QY 451 LeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465  
 Db 2802 CTCATAGGTATCAAAAGGAAGTACCTCTGTGACATCTTGGATG 2846

RESULT 7  
 BC046520 2925 bp mRNA linear PRI 21-FEB-2003  
 LOCUS

DEFINITION Homo sapiens, Similar to sorting nexin 14, clone IMAGE:5267454,  
 mRNA.  
 ACCESSION BC046520  
 VERSION BC046520.1 GI:28461364  
 KEYWORDS  
 SOURCE Homo sapiens (human)  
 ORGANISM  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 2925)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-FEB-2003) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov  
 COMMENT Contact: MGC help desk  
 Email: cgaaps-remail.nih.gov  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Baylor College of Medicine Human Genome  
 Sequencing Center  
 Center code: BCM-HGSC  
 Web site: http://www.hgsc.bcm.tmc.edu/cdna/  
 Contact: amg@bcm.tmc.edu  
 Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounsgood, H.,  
 Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
 A.N., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Series: TRAK Plate: 91 Row: a Column: 6.

## FEATURES

## source

1..2925  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5267454"  
 /tissue\_type="Testis"  
 /clone\_lib="NIH MGC 97"  
 /lab\_host="DH10B"  
 /note="Vector: pBluescript"

BASE COUNT 948 a 513 c 579 g 885 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 5,02e-197 Length: 2925  
 Score: 2371.00 Matches: 454  
 Percent Similarity: 99.7% Conservative: 0  
 Best Local Similarity: 99.7% Mismatches: 1  
 Query Match: 97.2% Indels: 0  
 DB: 9 Gaps: 0

US-09-744-313A-1 (1-465) x BC046520 (1-2925)

QY 11 ArgAsnThrGlnLysArgGlyCysSerPheGlyIleSerArgIleGlySerLysIleLys 30  
 Db 1381 AGGAACACACAGAAAAGGGAGAACTATTGGAAATCAGCAATAGTAGCAAAATTTAA 1440  
 QY 31 GlyValPheLysSerThrThrMetGluGlyAlaMetLeuProAsnTyrGlyValAlaGlu 50  
 Db 1441 GGAGTATTCAAAAGTATACCAATGAGGAGGTATGTTGCCTAATTATGCTAGCTGAA 1500  
 QY 51 GlyGluAspAspPheIleGluGlyIleValValMetGluAspAspSerProValGlu 70  
 Db 1501 GGTGAAGATGATTTTATTGAAGAAGGTAATTGTTGTAATGGAAGATGATTCTCCAGTGGAG 1560  
 QY 71 AlaValSerThrProAsnThrProArgAsnLeuAlaIleThrLysIleSerIleProTyr 90  
 Db 1561 GCTGTGACACACTTAATCTCCAGAAACCTTGCTGCATGGAAATAGCATTCCTAT 1620

451 LeuA5nlyValGlnlySgluValThrSerValThrSerTirpMet 465  
Db 2701 CTCAATAAGGTACAAAGGAAGTACCTCTGTGACATCTGGATG 2745

RESULT 8  
AK026479 3576 bp mRNA linear PRI 29-SEP-2000  
LOCUS Homo sapiens cDNA: FJ22826 f1s, clone KAI14022, highly similar to  
DEFINITION AF121863 Homo sapiens sorting nexin 14 (SNX14) mRNA.  
ACCESSION AK026479  
VERSION AK026479.1 GI:10439350  
KEYWORDS oligo capping; f1s (full insert sequence).  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,  
Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Tanigami,A.,  
Fujiwara,T., Ono,T., Yamada,K., Fujii,Y., Ozaki,K., Hirao,M.,  
Omori,Y., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T.,  
Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.  
TITLE NEDO human cDNA sequencing project  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 3576)  
AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,  
Shibahara,T., Tanaka,T. and Nakamura,Y.  
TITLE Direct Submission  
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,  
University of Tokyo, Laboratory of Genome Structure Analysis, Human  
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,  
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,  
Fax:81-3-5449-5416)  
COMMENT NEDO human cDNA sequencing project supported by Ministry of  
International Trade and Industry of Japan; cDNA full insert  
sequencing; Research Association for Biotechnology; cDNA library  
construction, 5'- & 3'-end one pass sequencing; Department of  
Virology and Human Genome Center, Institute of Medical Science,  
University of Tokyo (partly supported by Science and Technology  
Agency).

FEATURES  
source Location/Qualifiers  
1 . 3576  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="KAI14022"  
/tissue\_type="ileal mucosa"  
/clone\_lib="kai1"  
/note="Cloning vector pWE18SPL3"  
misc\_feature 1..3576  
/note="highly similar to AF121863 Homo sapiens sorting  
nexin 14 (SNX14) mRNA"  
BASE COUNT 1181 a 633 c 735 g 1027 t  
ORIGIN

Alignment Scores:  
Pred. No.: 7,91e-196 Length: 3576  
Score: 2358.50 Matches: 453  
Percent Similarity: 99.13% Conservative: 1  
Best Local Similarity: 98.91% Mismatches: 3  
Query Match: 96.78% Indels: 3  
DB: 9 Gaps: 1

US-09-744-313A-1 (1-465) x AK026479 (1-3576)

QY 11 ArgAenThrGlnlySgluSerPheGlyIleSerArgIleGlySerlyVallelys 30  
Db 2017 AGGACACACAGAAAGGGGAGATCATTTGGATTCAGACAGATGACCAATTAATA 2076  
QY 31 GlyValPheLySerThrThrMetGlyGlyAlaMetLeuProAenTyGlyValalaGlu 50  
Db 2077 GGAGTATTCAAAAGTACCAATGAGGAGACATGTTGCTTATTATGGTGAGTCAA 2130

QY 51 GlyGluAspPheIleGluGlyLeuValMetGluAspSerProValGlu 70  
DB 2137 GGTGAAGATGATTTATTTGAAGAAGGTATTTGTTGAATGAAGATGATTTCCAGTGGAG 2196  
QY 71 AlaValSerThrProAsnThrProArgAsnLeuAlaAlaTyrLysIleSerIleProTyr 90  
DB 2197 GCTGTGAGCACACCTTAATCTCCCGAACCCTTCTGTCATGGAATATAGCATTCATAT 2256  
QY 91 ValAspPheGluAspProSerSerGluArgLysGluLysGluArgIleProVal 110  
DB 2257 GTAGACTTTTTCAGGATCCCTCTCTGAAGAGGAGAGAAAGAAAGAAATTCCTGTG 2316  
QY 111 PheCysIleAspValGluArgAsnAspArgAsnAlaValGlyHisGluProGluHisTyr 130  
DB 2317 TTTTGTATTGATCTTGAAGAAATGATAGAGAGCAGTTCGACAGGCTGAAACATTCG 2376  
QY 131 SerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHis 150  
DB 2377 TCTGTCTATAGAGATATCTTGAATCTATGTAATCTTGAATCAAACTAACAAGTTTCAT 2436  
QY 151 GlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu 170  
DB 2437 GGTGCATTTCTGATGCCAGCTTCCTCTTAAGAGGATCATTCGGCCCCAAAATATGAA 2496  
QY 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190  
DB 2497 TTCTTAAAGTCAAGAGGAGGAGTTCAGAGATATCTCAGAAACTTCTGCAGCATCCA 2556  
QY 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
DB 2557 GAACGTGATGATATGTCAACTTCGTGGCAGATCTTCTTCCCTTAATGCTGGGAAACAA 2616  
QY 211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValProGly 230  
DB 2617 TTTTCTGATAGATATCTACAGATGTAATCTTGGGAAATATATAAACTGTTCTGGA 2676  
QY 231 LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPhelleAsnSer 250  
DB 2677 AAACATATGAAGAGAGAGTTCAGACTTCGGAACCTTTATCATGAATTCATTAATCT 2736  
QY 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
DB 2737 TGTGAGTCTCCAAAGCCTAAACCAAGTAGACCACTGACCACTTCAGCCCTCATCTCA 2796  
QY 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsn 290  
DB 2797 GAAACCAACACAGAGAGCTTTTCAAGTCTGTTTAAATAATATGCAACCCGTCGTAAT 2856  
QY 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
DB 2857 ACAGAGAGAAAGCAATATCAGATATTTTATGAGGTGATGACTGTAGAAGGAGTCTAT 2916  
QY 311 AspTyrLeuMetTyrValGlyArgValPheGlnValProAspTyrLeuHisLeu 330  
DB 2917 GATTACCTGATGATAGGACGGGTAGTTTCCAGGTTCCTGACTGGCTTCATCATCTC 2976  
QY 331 LeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
DB 2977 TTAATGGGAATCGGAATCCTCTTTAAACCAACCCCTGGAAATGATATCTGATTACTT 3036  
QY 351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeu 370  
DB 3037 CAGTGTAACTAGAACACACTATTTTCAGAGGACCGGTTGGTCTCTCATATAACACTTCTC 3096  
QY 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
DB 3097 AGAGATGCTATATTCTGTGAACCCCTGAACCTGCTCTCTCCAAAGATAAGCAAAAAGGA 3156  
QY 391 AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
DB 3157 GCAAAACACACTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGTATT 3216  
QY 411 GlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProVal 430

DB 3217 GGTGAAGAAACCCAGTATGAAGATCATCAGACTTCCTTTGATGGCTTACAGCAACCCAGTA 3276  
QY 431 LeuAsnLysGln-----LeuThrTyrValLeuLeuAspIleValIleGlnGluLeu 447  
DB 3277 CTCACCAAGCAGGTGCTTTAGCTGACTTATGTTTATTGACATTTGATACAGGAACCTG 3336  
QY 448 PheProGluLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
DB 3337 TTTCCAGAGCTCAATAGGTACAAAGGAAGTTACCTCTGTGACATCTTGGATG 3390  
RESULT 9  
AK095380 3616 bp mRNA linear PRI 15-JUL-2002  
LOCUS Homo sapiens cDNA FLJ38061 fis, clone CTONG2014966, highly similar  
DEFINITION to SORTING NEXIN 14.  
ACCESSION AK095380  
VERSION AK095380.1 GI:21754626  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M.,  
Shimizu F., Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K.,  
Sugiyama T., Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S.,  
Yamamoto J., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T.,  
Kimura K., Yamashita H., Mateo K., Nakamura Y., Sekine M.,  
Kikuchi H., Kanda K., Wagatsuma M., Murakawa K., Kanehori K.,  
Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y.,  
Sugano S., Nagahara K., Masuho Y., Nagai K. and Isogai T.  
NEBO human cDNA sequencing project  
Unpublished  
JOURNAL Isogai T. and Yamamoto J.  
REFERENCE 2 (bases 1 to 3616)  
AUTHORS Direct Submission  
JOURNAL Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
COMMENT NEBO human cDNA sequencing project supported by Ministry of  
Economy, Trade and Industry of Japan; cDNA full insert sequencing:  
Research Association for Biotechnology (RAB); cDNA library  
construction: Helix Research Institute (HRI) (supported by Japan  
Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,  
HRI, and Biotechnology Center, National Institute of Technology and  
Evaluation; clone selection for full insert sequencing: HRI and  
RAB; annotation: HRI and RAB.  
FEATURES  
source  
1. 3616  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CTONG2014966"  
/tissue\_type="tongue, tumor tissue"  
/clone\_lib="CTONG2"  
/note="Cloning vector: pME18SFL3"  
BASE COUNT 1078 a 629 c 710 g 1199 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1,98e-195 Length: 3616  
Score: 2354.00 Matches: 453  
Percent Similarity: 99.56% Conservative: 1  
Best Local Similarity: 99.34% Mismatches: 1  
Query Match: 96.59% Indels: 1  
DB: 9 Gaps: 0  
US-09-744-313A-1 (1-465) x AK095380 (1-3616)  
QY 11 ArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerLysIleLys 30  
DB 2080 AGGAACACACAGAAAAGGGGAGATCTTTGGAAATCAGCAATAGTAGCAAAATTAA 2139

QY 31 GlyValPheLysSerThrThrMetGluGlyValMetLeuProAsnTyrGlyValAlaGlu 50  
Db 2140 GGAAGTATTCAAAAGTACCACATGGAGGAGCTATGTTGCCCTAATTATGTTAGCTGAA 2199  
QY 51 GlyGluAspPheLeuGluGluGlyValValMetGluAspSerProValGlu 70  
Db 2200 GGTGAAGATGATTTATTGAAGAGGTATTGTTGTAATGGAAGATGATTCCTCAGTGGAG 2259  
QY 71 AlaValSerThrProAsnThrProArgAsnLeuAlaValTyrLysLeuSerLeuProTyr 90  
Db 2260 GCTGTGAGCACACCTAATACCTCCGAAACCTTCTGCTGATGGAAATTAAGCATTCATAT 2319  
QY 91 ValAspPheGluAspProSerSerGluArgGlyGlu-LysLysGluArgTleProVa 110  
Db 2320 GTAGACTTTTGTAGGATCCCTCTCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2379  
QY 110 lPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTr 130  
Db 2380 GTTTTGTATTGATTGATGTAAGAGATGATGATGAGAGGAGTGGACAGGCTGAAATG 2439  
QY 130 pSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheH 150  
Db 2440 GTCTGTCTATAGAAGATCTTGAATCTTATGATCTTGAATCAAACTAACAGAAATTC 2499  
QY 150 sGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGl 170  
Db 2500 TGGTGCATTTCTCTGATCCAGCTTCCTCTTAAGAGGATCATTTGGCCCCAAAATATGA 2559  
QY 170 uPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPr 190  
Db 2560 ATTTCTTAAGTCAAGAGGAGGAGGAGTCCAGAGATATCTACAGAACTTCTGAGGATCC 2619  
QY 190 oGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGl 210  
Db 2620 AGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2679  
QY 210 nPheLeuAspLysLysLeuProAspValAsnLeuGlyLysIleIleLysSerValProGl 230  
Db 2680 ATTTCTTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2739  
QY 230 YLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleLys 250  
Db 2740 AAAAATAAGTAAGAGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 2799  
QY 250 rCysGluSerProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
Db 2800 TTGTGAGTCTCAAGAGCTTAACCAAGTAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2859  
QY 270 rGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAs 290  
Db 2860 AGAAAAACAACAAGAGCTTTTCAATGATCTGTTTAAAAATAATGCAAAACCGTGTGAAA 2919  
QY 290 nThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTy 310  
Db 2920 TACAG 2979  
QY 310 rAspTyrLeuMetTyrValGlyArgValValPheGluValProAspThrLeuHisLys 330  
Db 2980 TGATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3039  
QY 330 uLeuMetGlyThrArgLysLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLe 350  
Db 3040 CTTAATGGGAAGTCTGATCTCTTTAAAAACACCTGGAAATGATATGATGATGATGATG 3099  
QY 350 uGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLe 370  
Db 3100 TCAGTGTAACTAGAACAGCTATTTTCAGGAGCAGCGTTTGGTCTCACTCAATAACACTCT 3159  
QY 370 uArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGl 390  
Db 3160 CAGAGATGCTATATTCTGTGAAACACTCAACCTCGCTCTCTCCCAAGATAGCAAAAGG 3219

QY 390 yAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIl 410  
Db 3220 AGCAAAACAGACTTTTGAAGAAATGATGATTAATCATTCAGATCTGTTAGTCAAGTGTAT 3279  
QY 410 eGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProVa 430  
Db 3280 TGGTGAAGAAACCAAGATGGAAGCATCAGACTTCTCTTTGATGGCTTTACAGCAACAGT 3339  
QY 430 lLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGl 450  
Db 3340 ACTCAACAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3399  
QY 450 uLeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465  
Db 3400 GCTCAATAAGGTACAAAGGAAGTTACCTCTGTGACATCTTGGATG 3445  
RESULT 10  
LOCUS AF121863 1593 bp mRNA linear PRI 14-SEP-2001  
DEFINITION Homo sapiens sorting nexin 14 (SNX14) mRNA, partial cds.  
ACCESSION AF121863  
VERSION AF121863.1 GI:4689265  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1593)  
AUTHORS Teasdale,R.D., Locci,D., Houghton,F., Karlsson,L. and Gleeson,P.A.  
TITLE A large family of endosome-localized proteins related to sorting nexin 1  
JOURNAL Biochem. J. 358 (Pt 1), 7-16 (2001)  
MEDLINE 21378165  
PUBMED 11485546  
REFERENCE 2 (bases 1 to 1593)  
AUTHORS Teasdale,R.D., Gleeson,P.A. and Karlsson,L.  
TITLE Identification of eleven novel human sorting nexin molecules. A sub-group of the sorting nexin family is associated with the early endosomes  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 1593)  
AUTHORS Teasdale,R.D., Gleeson,P.A. and Karlsson,L.  
TITLE Direct Submission  
JOURNAL Submitted (22-JAN-1999) The R.W. Johnson Pharmaceutical Research Institute, 3535 General Atomics Court, San Diego, CA 92121, USA  
FEATURES  
source  
1..1593  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
c1..1593  
/gene="SNX14"  
c1..1125  
/gene="SNX14"  
/codon\_start=1  
/product="sorting nexin 14"  
/protein\_id="AAD27836.1"  
/db\_xref="GI:4689266"  
/translation="DFRDPSSSRKKERIPVFCIDVERNDRAVGHPEHWSVYRR  
VLEFVYLSKLTTEHGAAPPDQLPSKRIIGPKNYEFLKSKEEFQVYLOKLOHPFLS  
NSQLADFLSPNGETFDKILFDVNLGKIKSVPGKLMKEQSHLEPFTMPTNSC  
ESPKPSPRLTILSPSTENNNKLFNDFKNNANRAETKONQNTFFMEVTVFEG  
YLLMYGVGVFQVDFMLHLHLMGTRILFKNTLEMTYDLYLQLEQLFOHRLVSLI  
TYLDRDAIPECTEPRSLQDKOKAKQTFEENMNTYIPDLLVKICIGETKYESIRLLFDG  
LQQPVNLKQLYVLLDIVIQELFPELNNKQKSVTSVTSM"

BASE COUNT 533 a 256 c 294 g 510 t

ORIGIN

Alignment Scores:

Pred. No.:	7,5e-162	Length:	1593
Score:	1964.00	Matches:	374
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0



```

432 AnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPheProGluLeu 451
1021 AACACGACGCTGACTATGTTTATTTGGACATGTGATACAGAACTGTTTCAGAGCTC 1080

452 AsnLysValGlnLysGluValThrSerValThrSerValThrMet 465
1081 AATAAGTACAAAGAGAGTTACCTCTGTGACATCTTGGATG 1122

RESULT 11
HSA420561      968 bp  mRNA  linear  PRI 23-NOV-2001
LOCUS      Homo sapiens mRNA full length insert cDNA clone EUROIMAGE 1521079.
DEFINITION
ACCESSION  AJ420561
VERSION    AJ420561.1  GI:17066425
KEYWORDS
SOURCE
ORGANISM    Homo sapiens (human)
REFERENCE
AUTHORS    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE      Auffray,C., Ansoerge,W., Ballabio,A., Estivill,X., Gibson,K.,
            Lehrach,H., Poustka,A. and Lundeberg,J.
            The European IMAGE consortium for integrated Molecular analysis of
            human gene transcripts
JOURNAL
AUTHORS
TITLE      Submitted (02-OCT-2001) Persson A., Center for Molecular
            Biotechnology, KTH, SCFAB, Institute of Biotechnology, Roslagsvagen
            30B, 106 91 Stockholm, SWEDEN
COMMENT    This clone is available royalty-free through IMAGE Consortium
            Distributors. IMPORTANT: This sequence represents the full insert
            of this IMAGE cDNA clone. No attempt has been made to verify
            whether this corresponds to the full-length of the original mRNA
            from which it was derived.
FEATURES
            source
            1..968
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /chromosome="6"
            /clone="EUROIMAGE 1521079"
            /clone_1lb="NCI CGAP Lu5"
            /clone_1lb="NCI CGAP Lu5"
            311 a 147 c 186 g _324 t

BASE COUNT  311 a 147 c 186 g _324 t
ORIGIN

Alignment Scores:
Pred No.:      1.09e-64      Length:      968
Score:         847.00      Matches:    163
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:   34.76%      Indels:    0
DB:            Gaps:      0

US-09-744-313A-1 (1-465) x HSA420561 (1-968)

303 ValMetThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGln 322
21 GTGATGACTGTAGAAGGAGTCTATGATTACCTGATGTATGTAGGACGGGTAGTTTCAG 80
323 ValProAspTyrPheUHisLeuLeuMetGlyThrGileLeuPheLysAsnThrLeu 342
81 GTTCCTGACTGGCTTCATCTCTTTATGGGAACTCGAATCCCTTTAAAAACACCCCTG 140
343 GluMetTyrThrAspTyrTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArg 362
141 GAAATGTATATCTGATTACTATCTTTCACTGTTAACTAGAACACGACTATTTCAGGAGCACCGT 200
363 LeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArg 382
201 TTGGTCTCACTCATCAACACTTCTCAGAGATGCTATTCTGTGAAAAACACTGAACCTCGC 260

```

Query Match:	80.59%	Indels:	0
DB:	9	Gaps:	0
US-09-744-313A-1 (1-465) x AF121863 (1-1593)			
QY	92	AspPheGluAspProSerSerGluArgLysGluLysLysGluArgLleProValphe	111
DB	1	GACTTTTGTAGCATCTCCCTCTCAAGGAAGGAGAAAAAGAANAATTCCTGTGTTTT	60
QY	112	CysIleAspValGluArgAsnAspArgAlaValGLyHISgluProGluHisTrpSer	131
DB	61	TGTATTGATGTTGAAAGAAATGATAGAGACGAGITGGACACGAGCCTGAACATTGGTCT	120
QY	132	ValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGly	151
DB	121	GTCTATAGAGATATCTTGAAATCTATGTACTTGAATCAAACTAACAGAAATTCATGGT	180
QY	152	AlaPheProAspAlaGlnLeuProSerLysArgIlelleGlyProLysAsnTyrGluPhe	171
DB	181	GCATTTCCGTGATGCCAGCTTCCTTCTTAAGAGGATCATGGCCCCCAAANAATTGAATTC	240
QY	172	LeuLysSerLysArgGluGluPheGlnGlnTyrLeuGlnLysLeuLeuGlnHisProGlu	191
DB	241	TTAAAGTCAAGAGGGAAGAGITCCDAGATAATCTACAGAACTTCTCGAGCATCCAGAA	300
QY	192	LeuSerAsnSerGlnLeuLeuAlaaspPheLeuSerProAsnGlyGlyGluThrGlnPhe	211
DB	301	CTGAGTAATAGTCAACTTCTGGCAGACTTCTTTCCCTTAATGGTGGGGAACAACAATTT	360
QY	212	LeuAspLysIleLeuProAspValLeuLeuGlyLysIlelleLysSerValProGlyLys	231
DB	361	CTTGATAAGANATCTACCAGATGAATCTTGGGAAAATTTATAAATCTGTGTTCTCGAANA	420
QY	232	LeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSerCys	251
DB	421	CTAATGAAGAGAAAGGTGAGCATTTGGAACTTTTATCATGAATTCATTAAATCTTGT	480
QY	252	GluSerProLysProLysProSerArgProGluLeuThrIleLysSerProThrSerGlu	271
DB	481	GAGTCTCCAAGGCTTAACCAAGTAGCACAGAACTGACCATTCTCAGCCCTACTTCAGAA	540
QY	272	AsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThr	291
DB	541	AACAACAAGAGCTTTTCAATGATCTGTGTTAAANAATATGCAACCGTGTCAANAATACA	600
QY	292	GluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAsp	311
DB	601	GAGAGAAAGCAAAATCAGAAATTTATTTATGGAGGTGATGACTGTAGAGAGTCTATGAT	660
QY	312	TyrLeuMetTyrValGlyArgValPheGlnValProAspTrpLeuHisIleLeuLeu	331
DB	661	TACCTGATGTATGTAGACGGGTAGTTTTCCAGGTTCTCGACGGTTCATCATCTCTTTA	720
QY	332	MetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGln	351
DB	721	ATGGGAACCTGAACTCTCTTTAAACAACCTCGMAATGTATACTGATTACTATCTCTAG	780
QY	352	CysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeulleThrLeuLeuArg	371
DB	781	TGTAACCTAGAACAGCTATTTCAGAGCACCGTTTGGTCTCTCACTATAACACTTCTCAGA	840
QY	372	AspAlaIlePheCysGluIleAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAla	391
DB	841	GATGCTATATTCTGTGAAACACTGAACCTCGTCTCTCCAAATAGCAAAAGAGACA	900
QY	392	LysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGly	411
DB	901	AAAACAGACTTTTGAAGAAATGATGAATTAACAATCCAGATCTGTGTAGTCAAGTGTATGGT	960
QY	412	GluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnProValLeu	431
DB	961	GAAGAAACAAGTATGAAGCATCAGACTTCTGTGTTTGTGGCTTACAGCAACAGTACTC	1020

383 SerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIle 402  
Db 261 TCTCTCCAGATAGCAAAAGAGCAAAACAGACTTTTGAAGAAATGATGATACATT 320  
403 ProAspLeuLeuValLysCysIleGlyGluGlnThrLysTyrGluSerIleArgLeuLeu 422  
Db 321 CCAGATCTGTAGTCAAGTCTATTGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCG 380  
423 PheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIle 442  
Db 381 TTTGATGGCTTACAGCAACCAAGTACTCAACAGCAGCTGACTTATGTTTTATTGGACATT 440  
443 ValIleGlnGluLeuPheProGluLeuAsnLysValGlnLysGlnValThrSerValThr 462  
Db 441 GTGATCAGAACTGTTTCCAGAGCTCAATPAGGTACAAAGGAAGTTTACCTCTGGACA 500  
463 SerTyrMet 465  
Db 501 TCTTGGATG 509

RESULT 12  
BD113855  
LOCUS BD113855 425 bp DNA linear PAT 18-SEP-2002  
DEFINITION EST and encoded human protein.  
ACCESSION BD113855  
VERSION BD113855.1 GI:23208759  
KEYWORDS JP 2002010789-A/5932.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 425)  
AUTHORS Edwards,J.B.D.M., Jobert,S. and Giordano,J.E.  
TITLE EST and encoded human protein  
JOURNAL Patent: JP 2002010789-A 5932 15-JAN-2002;  
GENSET CORP  
COMMENT OS Homo sapiens (human)  
PN JP 2002010789-A/5932  
PD 15-JAN-2002  
PF 07-AUG-2000 JP 2000280989  
PR 05-AUG-1999 US 60/147499  
PI JEAN BAPUIST DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EVE  
GIORDANO  
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC  
C12N15/00  
CC EST and encoded human protein  
FH Key Location/Qualifiers  
FT source 1..425  
/organism="Homo sapiens (human)"

FEATURES  
source  
1..425  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 134 a 82 c 85 g 124 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 4.62e-56 Length: 425  
Score: 743.00 Matches: 141  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 30.49% Indels: 0  
DB: 6 Gaps: 0

US-09-744-313A-1 (1-465) x BD113855 (1-425)  
Qy 297 GlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrVal 316  
Db 1 CAGAAATTTATGAGGAGTATGACTGTAGAGGAGTCTATGATTACCTGATGATGTA 60

317 GlyArgValValPheGlnValProAspTyrPheLeuHisLeuLeuMetGlyThrArgIle 336  
Db 61 GGACGGGTAGTTTCCAGGTTCTGACTGGCTTCATCATCTCTTAATGGGAATCGAATC 120  
337 LeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGln 356  
Db 121 CTTCTTAAACACACCCCTGGAAATGTATATCTGATTACTATCTTCACTGCTAAACTAGAACAG 180  
357 LeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCys 376  
Db 181 CTATTTCCAGGAGCACCCCTTTGGTCTCTCACTCATACACTTCTCAGAGATGTATATTCTGT 240  
377 GluAsnThrGluProAspSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGlu 396  
Db 241 GAAACACACTGAACTCTCTCTCCAGATAAGCAAAAGAGCAAAACAGACTTTTGAA 300  
397 GluMetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGlnThrLysTyr 416  
Db 301 GAAATGATGATTAATTCAGATCTGTAGTCAAGTGTATTGGTGAAGAAACCAAGTAT 360  
417 GluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThr 436  
Db 361 GAAAGCATCAGACTTCTGTTTGTATGAGTTCAGCAACCAAGTACTCAACAGAGCTGACT 420  
437 Tyr 437  
Db 421 TAT 423

RESULT 13  
BD026500  
LOCUS BD026500 358 bp DNA linear PAT 27-AUG-2002  
DEFINITION Sequence tag and encoded human protein.  
ACCESSION BD026500  
VERSION BD026500.1 GI:22567723  
KEYWORDS JP 2001269182-A/2746.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 358)  
AUTHORS Edwards,J.B.D.M., Duclair,G. and Jordan,J.Y.  
TITLE Sequence tag and encoded human protein  
JOURNAL Patent: JP 2001269182-A 2746 02-OCT-2001;  
GENSET  
COMMENT OS Homo sapiens (human)  
PN JP 2001269182-A/2746  
PD 02-OCT-2001  
PF 24-FEB-2000 JP 2000118773  
PR 26-FEB-1999 US 60/122487  
PI JEAN BAPUIST DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES  
JORDAN  
PC C12N15/09, C07K14/435, C07K16/18, C12N1/15, C12N1/19, C12N1/21, PC  
C12N5/10,  
PC C12P21/02, C12P21/08, C12Q1/68, C06F17/30, C12N15/00, C12N5/00, PC  
G06F15/40  
CC  
FH Key Location/Qualifiers  
FT CDS 53..358.  
Location/Qualifiers  
1..358  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 125 a 63 c 69 g 101 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1.2e-46 Length: 358  
Score: 634.00 Matches: 119  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 26.02% Indels: 0  
DB: 6 Gaps: 0

US-09-744-313A-1 (1-465) x BD026500 (1-358)

Qy 216 LeuProaspValAsnLeuGlyLysIlelleLysSerValProGlyLysLeuMetLysGlu 235  
 Db 2 CTACCAGATGAATCTTGGGAAATATATAATCTGCTCGGAAACATAATGAAGAG 61  
 Qy 236 LysGlyGlnHisLeuGluProPheIleMetAsnPhelleAsnSerCysGluSerProLys 255  
 Db 62 AAAGGTGAGCATTTGGAAACCTTTTATCATGAATTTTCAATTAATCTTGTGAGTCTCAAG 121  
 Qy 256 ProLysProSerArgProGluLeuThrIleLeuSerProThrSerGluAsnLeuLysLys 275  
 Db 122 CCTAAACCAAGTAGACCAAGACTGACCATCTCAGCCCTACTTCAGAAACACACAGAAG 181  
 Qy 276 LeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGln 295  
 Db 182 CTITTCATGATCTGTTTAAATAATATGCAACCGTGTGAAATATACAGAGAAAGCAA 241  
 Qy 296 AsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyr 315  
 Db 242 AATCAGAAATATTTTATGAGAGTGATGACTGTAGAGGAGCTCATGATTAACCTGATGTAT 301  
 Qy 316 ValGlyArgValPheGlnValProAspTyrLeuHisLeuMetGlyThr 334  
 Db 302 GTAGGACGGGTAGTCTTCCAGGTCTCTGACTGCTTCATCATCTCTTAATGGGAACC 358

RESULT 14  
 AC116713 278898 bp DNA linear HTG 09-JUN-2003  
 LOCUS Mus musculus clone RP23-118M23, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 7  
 DEFINITION ordered pieces.

AC116713 4 GI:31544102  
 HTG; HTGS PHASE2; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.

# ORGANISM

Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 278898)  
 Birren, B., Nusbaum, C. and Lander, E.  
 Mus musculus, clone RP23-118M23  
 Unpublished

# REFERENCE

1 (bases 1 to 278898)  
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,  
 Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,  
 Boukhgalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,  
 Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,  
 Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,  
 Fero, S., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S.,  
 Ginde, S., Gord, S., Govette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kellis, C., Larocque, K., Lamazares, R.,  
 Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,  
 MacLean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,  
 McCarthy, M., McSwan, P., McKernan, K., Meldrim, J., Meneus, L.,  
 Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,  
 Norou, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D.,  
 Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,  
 Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,  
 Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schuback, R.,  
 Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Straus, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
 Viel, R., Vo, A., Wilson, D., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
 Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (02-APR-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 278898)  
 Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N.,  
 Anderson, S., Arachchi, H. M., Barna, N., Bastien, V., Bloom, T.,  
 Boguslavskiy, L., Boukhgalter, B., Camarata, J., Chang, J., Choepel, Y.,

Collymore, A., Cook, A., Cooke, P., Corum, B., Dearellano, K.,  
 Diaz, J. S., Dodge, S., Dooley, K., Dorris, L., Erickson, J., Fero, S.,  
 Ferreira, P., FitzGerald, M., Gage, D., Galagan, J., Gardyna, S.,  
 Graham, L., Grand-Pierre, N., Hagos, B., Hagopian, D., Hagos, B.,  
 Hall, J., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
 Kamat, A., Karatas, A., Kellis, C., Landers, T., Levine, R.,  
 Lindblad-Toh, K., Liu, G., Liu, A., Mabbitt, R., MacLean, C.,  
 Macdonald, P., Major, J., Manning, J., Matthews, C., McCarthy, M.,  
 Meldrim, J., Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J.,  
 Nguyen, C., Nicol, R., Norou, C., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,  
 Rachupka, A., Ramasamy, U., Raymond, C., Retta, R., Rise, C., Rogov, P.,  
 Roman, J., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C.,  
 Spencer, B., Stange-Thomann, N., Stojanovic, N., Stubbs, M.,  
 Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M.,  
 Vassiliev, H., Venkataraman, V. S., Viel, R., Vo, A., Wilson, B., Wu, X.,  
 Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
 Direct Submission  
 Submitted (09-JUN-2003) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Jun 9, 2003 this sequence version replaced gi:29171427.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence\_submission@genome.wi.mit.edu

----- Project Information

Center project name: L22409

Center clone name: 118\_M\_23

-----

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 7 contigs. Gaps between the contigs  
 \* are represented as runs of N. The order of the pieces  
 \* is believed to be correct as given, however the sizes  
 \* of the gaps between them are based on estimates that have  
 \* been provided by the submitter.  
 \* This sequence will be replaced  
 \* by the finished sequence as soon as it is available and  
 \* the accession number will be preserved.

\* the accession number will be preserved.  
 \* 1 97014 97013: contig of 97013 bp in length  
 \* 97014 97113: gap of 100 bp  
 \* 97114 144931: contig of 47818 bp in length  
 \* 144932 145031: gap of 100 bp  
 \* 145032 171184: contig of 26153 bp in length  
 \* 171185 171284: gap of 100 bp  
 \* 171285 221380: contig of 50096 bp in length  
 \* 221381 221480: gap of 100 bp  
 \* 221481 224451: contig of 2371 bp in length  
 \* 224452 249724: contig of 25173 bp in length  
 \* 249725 249824: gap of 100 bp  
 \* 249825 278898: contig of 29074 bp in length.

# FEATURES

source  
 1..278898  
 /organism="Mus musculus"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:10090"  
 /clone="RP23-118M23"  
 /clone\_lib="RPC1-23 Female Mouse BAC"  
 BASE COUNT 84330 a 58884 c 57710 g 76354 t 1620 others  
 ORIGIN

# Alignment Scores:

Pred. No.: 1.74e-38 Length: 278898  
 Score: 581.00 Matches: 160  
 Percent Similarity: 44.12% Conservative: 20  
 Best Local Similarity: 39.22% Mismatches: 59  
 Query Match: 23.84% Indels: 170  
 DB: 2 Gaps: 8

172 LeuLysSerLysArgGluGluPheGlnGluThrLeuGlnLysLeuLeuGlnHisProGlu 191  
 226685 CTCGGAGCCTGAGGAGGAGGTTGTGAGTTTGGAGCCAGCTGGAGCTACATAGCTTT 226626  
 192 LeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGlnPhe 211  
 226625 GCCACGCTCTGCATCTCACCGGAAAAAACAATAAATGAGGGGGCAGAGATGCAACTT 226566  
 212 LeuAspLysIleLeuProAspValAsnLeuGly-----Lys 223  
 226565 AGTGATAGGATGTTGGCTGGCTTACTTGTAGGCTCTGGATAGTTAGAAATATCAACG 226506  
 224 IleIleLysSerValProGlyLysLeuMetLysGluGlyGlnHisLeuGluProPhe 243  
 226505 ACCATAAGCGACACCAACAAAAATGGTGAATAA-CAAGAAAGAAAAAGAGATGCTTT 226447  
 244 -----IleMetAsnPheIleAsnSerCys----- 251  
 226446 ATGGCTAGTAATAATTAAGTTATGGCAAAAGTGAATCAACGCTTTTATTAAATGTTTGA 226387  
 252 -----GluSerProLysPro 256  
 226386 GAGAAATATGGGATTTTTTTTTTATAGCTTTGGTCATGAATGAGAAAGTTTGAAGCA 226327  
 257 LysProSerArgProGlu-----LeuThrIleLeuSerProThr 269  
 226326 GATTTTAACTTCTTAATGTGGCATTTTGTGATTTCTTTTAACTTATGATGATCTGTC 226267  
 270 SerGluAsnLysLys-----LeuPheAsn 278  
 226266 AATAATGCAATAAATCTGTCTCTTTATTATTTCACTTATATCTTACGTTAGTTTCAAC 226207  
 279 AspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsn 298  
 226206 GATCTGTTTAAAGATAATGCAACCGCGCTGAGAACACAGAGGAAAGCAAAATCAGAAC 226147  
 299 TyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyrVal----- 316  
 226146 TACTTCAGGAGGTGATGACTGTGGAGGAGTCTATGATTTACCTGATGATGATGATGTAAG 226087  
 316 ----- 316  
 226086 GCTTAATGGTCCAGTTGTTCTTTTGTAGTCTGCTCATGATTCAGATACAGACCTGCTC 226027  
 316 ----- 316  
 226026 AACTTTGGAAGAAATTTCTTTCCCGATTTACTACATTTGGCTTGACATGACCACTG 225967  
 316 ----- 316  
 225966 TTACTTTTATTCATGTATCATCTTGAGTGACACAGATTTCACTTCAATTTGATGCTC 225907  
 317 -----GlyArgValValPheGlnValProAspTyrLeuHisLeuLeuMetGlyTh 334  
 225906 TTCCTCAGGCGAGTAGTTTTCAGTCCAGACTGCTTCATCACTTCTTATGGAAC 225847  
 334 rArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeuGlnCysLysle 354  
 225846 TCGAATCTCTTTAAGAACACCTCGGAAATGTACACAGACTTACTCTCAGTGCAGCT 225787  
 354 uGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeuArg----- 371  
 225786 GGAGCAGCTGTTTCAGGAGCACCGGCTGGTCTCCTCATCACCCTTCTCAGAGTTTCGTA 225727  
 371 ----- 371  
 225726 TGACCATGCTGTATCATGGTCTACCTTTGTGTTTTTTTTTTTCCAGGATTTTAAACC 225667  
 371 ----- 371  
 225666 AGCATTTTATTACTAGATATACTTGGTATAAATTGGAGAAATATGTATATATTC 225607

372 -----AspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspL 387  
 225606 TGTACTCATTTTGTAGATGCTATATTTGTGAAATATCTGAACCTCGCTCTCCAGATA 225547  
 387 YSGlnLysGlyValAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuV 407  
 225546 AGCAAAAGGAGCAAAACAGACTTTTGAGAAATGATGAATTATCTCCAGGTATGATAT 225487  
 407 allYs-----CysIle 410  
 225486 TTAATAACTCAATGTGATA 225467

RESULT 15  
 AL589666/c  
 LOCUS  
 DEFINITION Human DNA sequence from clone RP11-321N4 on chromosome 6, complete sequence.  
 ACCESSION AL589666  
 VERSION AL589666.5 GI:14041764  
 KEYWORDS HTG.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 129010)  
 Direct Submission  
 Submitted (12-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On May 15, 2001 this sequence version replaced gi:13751565.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormPep; Information on the WormPep database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RP11-321N4 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm> VECTOR: pBAC3.6  
 IMPORTANT: This sequence is not the entire insert of clone RP11-321N4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP1-3U17 is at 128911 in this sequence. The true right end of clone RP11-30P6 is at 100 in this sequence.  
 This location/Qualifiers  
 1. 129010  
 /organism="Homo sapiens"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:9606"  
 /chromosome="6"  
 /clone="RP11-321N4"  
 /clone\_lib="RPCI-11.2"  
 2. 192  
 /note="MER21B repeat: matches 3. 207 of consensus"  
 repeat\_region

FEATURES  
 source

```
repeat_region 240. .909 /note="MER21B repeat: matches 30. .715 of consensus"  
repeat_region 1593. .1682 /note="MIR repeat: matches 35. .145 of consensus"  
repeat_region 2030. .2102 /note="L2 repeat: matches 2681. .2750 of consensus"  
repeat_region 3459. .3574 /note="L2 repeat: matches 2681. .2750 of consensus"  
repeat_region 5635. .5670 /note="FLAM_C repeat: matches 1. .117 of consensus"  
repeat_region 6887. .6982 /note="LIME repeat: matches 683. .718 of consensus"  
repeat_region 8284. .8356 /note="L2 repeat: matches 2616. .2704 of consensus"  
repeat_region 8380. .8405 /note="L2 repeat: matches 2639. .2710 of consensus"  
repeat_region 8440. .8601 /note="L3 copies 2 mer tt 100% conserved"  
repeat_region 8825. .9360 /note="MIR repeat: matches 1. .169 of consensus"  
repeat_region 9423. .9592 /note="L2 repeat: matches 1686. .2301 of consensus"  
repeat_region 12088. .12175 /note="L2 repeat: matches 2560. .2727 of consensus"  
repeat_region 12110. .12177 /note="44 copies 2 mer at 73% conserved"  
repeat_region 12790. .12905 /note="L7 copies 4 mer atat 79% conserved"  
repeat_region 13672. .13923 /note="L2 repeat: matches 2628. .2750 of consensus"  
repeat_region 14774. .15053 /note="L2 repeat: matches 2462. .2725 of consensus"  
repeat_region 15966. .16031 /note="Alusq repeat: matches 1. .289 of consensus"  
repeat_region 17304. .17359 /note="6 copies 11 mer 77% conserved"  
repeat_region 17554. .17682 /note="MIR repeat: matches 196. .251 of consensus"  
repeat_region 18366. .18701 /note="MIR repeat: matches 108. .228 of consensus"  
repeat_region 18959. .19248 /note="MER2 repeat: matches 1. .345 of consensus"  
repeat_region 20386. .20592 /note="AluJo repeat: matches 1. .284 of consensus"  
repeat_region 20609. .20778 /note="L2 repeat: matches 70. .288 of consensus"  
repeat_region 20747. .21117 /note="LTR28 repeat: matches 848. .1013 of consensus"  
repeat_region 21163. .21226 /note="LTR1 repeat: matches 295. .680 of consensus"  
repeat_region 21187. .21359 /note="LTR28 repeat: matches 297. .361 of consensus"  
repeat_region 22191. .22355 /note="LTR1 repeat: matches 16. .188 of consensus"  
repeat_region 22844. .22991 /note="FRAM repeat: matches 4. .163 of consensus"  
repeat_region 23155. .23190 /note="MIR repeat: matches 48. .202 of consensus"  
repeat_region 24390. .24440 /note="L8 copies 2 mer ga 86% conserved"  
repeat_region 25082. .25109 /note="L2 repeat: matches 2657. .2708 of consensus"  
repeat_region 25438. .25744 /note="7 copies 4 mer aac 100% conserved"  
repeat_region 25731. .25917 /note="L2 repeat: matches 2427. .2744 of consensus"  
repeat_region 25964. .26307 /note="MIR repeat: matches 95. .262 of consensus"  
repeat_region 26526. .26673 /note="L2 repeat: matches 1986. .2366 of consensus"  
repeat_region 26682. .26737 /note="FLAM_C repeat: matches 1. .142 of consensus"  
repeat_region 28208. .28394 /note="14 copies 4 mer agga 76% conserved"
```

```
repeat_region 29052. .29250 /note="MIR repeat: matches 12. .196 of consensus"  
repeat_region 30082. .30170 /note="MIR repeat: matches 48. .251 of consensus"  
repeat_region 30197. .31090 /note="MIR repeat: matches 50. .145 of consensus"  
repeat_region 31921. .32067 /note="MER11D repeat: matches 1. .897 of consensus"  
repeat_region 32545. .32678 /note="FRAM repeat: matches -2. .151 of consensus"  
repeat_region 32860. .32895 /note="L2 repeat: matches 2578. .2708 of consensus"  
repeat_region 32897. .33193 /note="9 copies 4 mer ttta 83% conserved"  
repeat_region 35445. .35478 /note="Alusx repeat: matches 1. .296 of consensus"  
repeat_region 37442. .37598 /note="MIR repeat: matches 84. .118 of consensus"  
repeat_region 38659. .38860 /note="MER5B repeat: matches 1. .173 of consensus"  
repeat_region 40552. .40911 /note="MIR repeat: matches 15. .236 of consensus"  
repeat_region 40912. .42540 /note="THE1B repeat: matches 1. .364 of consensus"  
repeat_region 44641. .45020 /note="THE1B-INTERNAL repeat: matches 1. .1580 of consensus"  
repeat_region 45689. .45988 /note="MIR repeat: matches 1. .382 of consensus"  
repeat_region 46220. .46351 /note="Alusg repeat: matches 1. .302 of consensus"  
repeat_region 46358. .46609 /note="Tiger3b repeat: matches 567. .698 of consensus"  
repeat_region 46623. .46758 /note="Aluub repeat: matches 40. .290 of consensus"  
repeat_region 46759. .47229 /note="AluJo/FRAM repeat: matches 170. .301 of consensus"  
repeat_region 47230. .47442 /note="Tiger3b repeat: matches 2. .540 of consensus"  
repeat_region 47715. .48005 /note="MER44A repeat: matches 85. .327 of consensus"  
repeat_region 48577. .48973 /note="L1PA10 repeat: matches 5864. .6153 of consensus"  
repeat_region 50633. .50925 /note="MIR repeat: matches 1. .390 of consensus"  
repeat_region 53828. .55056 /note="Aluy repeat: matches 3. .295 of consensus"  
repeat_region 55092. .56081 /note="L1MC1 repeat: matches 5023. .6332 of consensus"  
repeat_region 56074. .56651 /note="L1M4 repeat: matches 3778. .4798 of consensus"  
repeat_region 56652. .56952 /note="L1MCA repeat: matches 579. .1150 of consensus"  
repeat_region 56953. .57277 /note="Alusx repeat: matches 1. .301 of consensus"  
repeat_region 57280. .57562 /note="L1MCA repeat: matches 253. .579 of consensus"  
repeat_region 57699. .57766 /note="L2 repeat: matches 1590. .1886 of consensus"  
repeat_region 59295. .59599 /note="L2 repeat: matches 2639. .2710 of consensus"  
repeat_region 59749. .59811 /note="Alusx repeat: matches 1. .307 of consensus"
```

Alignment Scores:  
Pred. No.:  
Score:

1.24e-38 Length:  
578.00 Matches:  
129010  
129

Percent Similarity: 46.45% Conservative: 2  
Best Local Similarity: 45.74% Mismatches: 2  
Query Match: 23.72% Indels: 149  
DB: Gaps: 2

US-09-744-313A-1 (1-465) x AL589666 (1-129010)

Qy 276 LeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGluAsnThrGluArgLysGln 295  
Db 58941 CTTTTCATCATCTGTTTAAATAATGCAAAACCGTCTGAAATAACAGAGAGAAAGCAA 58882

Qy 296 AsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyrAspTyrLeuMetTyr 315  
Db 58881 AATCAGAATATTATTTATGGAGGTGATGCTGTAAGAGGAGTCTATGATTACCTGATGAT 58822

Qy 316 Val----- 316  
Db 58821 GTAGTAAAGATCTAAGTGGTTAAGACTGTTTTTCATTCATCTTACTTCTCTGTGATG 58762

Qy 316 ----- 316  
Db 58761 CAAATATGATCTTGATGATTGCCCATCATACTGCTCTTTTAAATTGAAGATACCTTTT 58702

Qy 316 ----- 316  
Db 58701 ATAATAATTCATACGATTGCCCTTGATCATATTAAGTTGCTAGTGTTCACAAATAAAA 58642

Qy 316 ----- 316  
Db 58641 TATTGACCATGGTCATTTGTGTTTTCTCATTTGGCAAAATTCAAAAATACCAATTCATCT 58582

Qy 317 ----- 317  
Db 58581 GGAACAATGAGGTATCCCATTTGATCTCTCTCAGACGGGTAGTTTTCCAGGTTCTCTGA 58522

Qy 325 pTTPLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyr 345  
Db 58521 CTGGCTTCATCACTCTTTAATGGGAAGTCTGAACTCTCTTTTAAACACACCTGGAAATGA 58462

Qy 345 rThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSe 365  
Db 58461 TACTGATTACTATCTTCAGTGTAACTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTC 58402

Qy 365 rLeuIleThrLeuLeuArg----- 371  
Db 58401 ACTATAACACTTCTCAGAGGTTTGATTTTCTAAATGTGTTGTTTCATATTTTGTGTAC 58342

Qy 371 ----- 371  
Db 58341 TCTATGATTTTAAATGATCTAAGAGAGAAATCTTGTGTTTGTCTTGTGTTTATTAGTA 58282

Qy 371 ----- 371  
Db 58281 TTTCTACTATGAGGAATCCTTTGTGTATAATAGATAAAATTCAGAAATATTTCTTT 58222

Qy 372 ----- 372  
Db 58221 TTACTCTTTTATGATCTATTTCTGTGAAACACACTGAACCTGCTCTCTCCAGATAA 58162

Qy 387 sGlnLysGlyValLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuVa 407  
Db 58161 GCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAAATACATTCACGCTGTAATATT 58102

Qy 407 Lys 408  
Db 58101 TAAG 58098

Search completed: January 31, 2004, 04:23:04  
Job time : 5430 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 02:52:57 ; Search time 497 Seconds  
(without alignments)

3409.897 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 2437

Sequence: 1 MYLHFLIPRNTKRGESF.....ELFPELNKVKQEVTSVTSMW 465

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 2434939 seqs, 1822278265 residues

Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO.spool/US09744313/runat\_30012004\_113538\_4278/app\_query.fasta\_1.647  
-DS=Published Applications NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1  
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100  
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcr -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -USER=US09744313 @CNG 1.1 271 @runat\_30012004\_113538\_4278  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications NA:  
1: /cgn2\_6/ptodata/2/pubpna/US07\_PUBCOMB.seq:  
2: /cgn2\_6/ptodata/2/pubpna/PCT\_NEW\_PUB.seq:  
3: /cgn2\_6/ptodata/2/pubpna/US06\_NEW\_PUB.seq:  
4: /cgn2\_6/ptodata/2/pubpna/US06\_PUBCOMB.seq:  
5: /cgn2\_6/ptodata/2/pubpna/US07\_NEW\_PUB.seq:  
6: /cgn2\_6/ptodata/2/pubpna/PCTUS\_PUBCOMB.seq:  
7: /cgn2\_6/ptodata/2/pubpna/US08\_NEW\_PUB.seq:  
8: /cgn2\_6/ptodata/2/pubpna/US08\_PUBCOMB.seq:  
9: /cgn2\_6/ptodata/2/pubpna/US09A\_PUBCOMB.seq:  
10: /cgn2\_6/ptodata/2/pubpna/US09B\_PUBCOMB.seq:  
11: /cgn2\_6/ptodata/2/pubpna/US09C\_PUBCOMB.seq:  
12: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
13: /cgn2\_6/ptodata/2/pubpna/US09\_NEW\_PUB.seq:  
14: /cgn2\_6/ptodata/2/pubpna/US10A\_PUBCOMB.seq:  
15: /cgn2\_6/ptodata/2/pubpna/US10B\_PUBCOMB.seq:  
16: /cgn2\_6/ptodata/2/pubpna/US10\_NEW\_PUB.seq:  
17: /cgn2\_6/ptodata/2/pubpna/US60\_NEW\_PUB.seq:  
18: /cgn2\_6/ptodata/2/pubpna/US60\_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
------------	-------	--------------------	-------	-------------

1	2386	97.9	1716	13	US-10-117-722-922	Sequence 922, App
2	2386	97.9	1716	15	US-10-037-270-922	Sequence 922, App
3	2386	97.9	3470	15	US-10-198-846-13016	Sequence 13016, A
4	2057.5	84.4	1551	13	US-10-117-722-923	Sequence 923, App
5	2057.5	84.4	1551	15	US-10-037-270-923	Sequence 923, App
6	981.5	40.3	725	9	US-09-910-943-712	Sequence 712, App
7	687	28.2	397	11	US-09-918-995-5183	Sequence 5183, App
8	682	28.0	402	10	US-09-983-965-540	Sequence 540, App
9	651	26.7	471	12	US-10-242-535A-25415	Sequence 25415, A
10	544	22.3	508	11	US-09-918-995-20205	Sequence 20205, A
11	505	20.7	305	10	US-09-983-965-355	Sequence 355, App
12	488	20.0	446	12	US-10-242-535A-49594	Sequence 49594, A
13	282	10.8	2720	12	US-10-108-260A-839	Sequence 839, App
14	249.5	10.2	2494	11	US-09-822-846-238	Sequence 238, App
15	235	10.2	280	10	US-09-783-590-4866	Sequence 4866, App
16	235	9.6	2456	10	US-09-764-868-437	Sequence 437, App
17	197	8.1	535	11	US-09-918-995-22530	Sequence 22530, A
18	160.5	6.6	384	10	US-09-783-590-4845	Sequence 4845, App
19	155.5	6.4	2391	12	US-10-295-027-115	Sequence 115, App
20	155.5	6.4	2760	10	US-10-098-841-195	Sequence 195, App
21	152.5	6.3	1294	14	US-09-764-868-198	Sequence 198, App
22	149.5	6.1	2512	9	US-09-784-249-1	Sequence 1, Appl
23	147.5	6.1	1502	9	US-09-822-849A-28	Sequence 28, Appl
24	147.5	6.1	3106	13	US-09-814-353-21324	Sequence 21324, A
25	142	5.8	447	11	US-09-918-995-35777	Sequence 35777, A
26	142	5.8	457	11	US-09-918-995-12082	Sequence 12082, A
27	137.5	5.6	1507	10	US-09-764-868-438	Sequence 438, App
28	137	5.6	1513	9	US-09-925-302-176	Sequence 176, App
29	136.5	5.6	1497	13	US-10-117-722-1021	Sequence 1021, App
30	136.5	5.6	1497	15	US-10-037-270-1021	Sequence 1021, App
31	136.5	5.6	2058	13	US-09-971-392-31	Sequence 31, Appl
32	132	5.4	2004	13	US-09-971-392-30	Sequence 30, Appl
33	128	5.3	1974	13	US-09-873-319-633	Sequence 633, App
34	128	5.3	1974	13	US-09-960-706-971	Sequence 434, App
35	126	5.2	1040	10	US-09-764-868-434	Sequence 316, App
36	126	5.2	1732	10	US-09-864-864-316	Sequence 671, App
37	125	5.1	1718	15	US-10-106-698-671	Sequence 25358, A
38	124	5.1	2628	12	US-10-369-493-25358	Sequence 1, Appl
39	123	5.0	1210	10	US-09-808-701-1	Sequence 1, Appl
40	123	5.0	1210	12	US-10-240-145-27	Sequence 1, Appl
41	123	5.0	1210	15	US-10-233-131-1	Sequence 1, Appl
42	119.5	4.9	1332	12	US-10-369-493-25530	Sequence 25530, A
43	119.5	4.9	3421	13	US-09-814-353-19872	Sequence 19872, A
44	118.5	4.9	3321	13	US-10-032-585-6365	Sequence 6365, App
45	117.5	4.8	418550	12	US-10-292-798-1463	Sequence 1463, App

#### ALIGNMENTS

#### RESULT 1

US-10-117-722-922  
Sequence 922, Application US/10117722  
Publication No. US20030219744A1  
GENERAL INFORMATION:  
APPLICANT: Tang, Y. Tom  
APPLICANT: Liu, Chenghua  
APPLICANT: Asundi, Vinod  
APPLICANT: Zhang, Jie  
APPLICANT: Drmanac, Radoje T.  
TITLE OF INVENTION: NO. US20030219744A1el Nucleic Acids and  
TITLE OF INVENTION: Polyptides  
FILE REFERENCE: 784CIP2BCIP  
CURRENT APPLICATION NUMBER: US/10/117,722  
CURRENT FILING DATE: 2002-04-04  
PRIOR APPLICATION NUMBER: 09/620,312  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 09/552,317  
PRIOR FILING DATE: 2000-04-25  
PRIOR APPLICATION NUMBER: 09/488,725  
PRIOR FILING DATE: 2000-01-21  
NUMBER OF SEQ ID NOS: 1104  
SOFTWARE: Pt\_FL\_genes Version 1.0  
SEQ ID NO 922

```

; LENGTH: 1716
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (71)..(1543)
US-10-117-722-922

Alignment Scores:
Pred. No.: 1716
Score: 2386.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 97.91%
Indels: 0
Gaps: 13

US-09-744-313a-1 (1-465) x US-10-117-722-922 (1-1716)

QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerIle 29
Db 173 TTTTCGGAACACACAGAAAGGGAGATCATTTGGATCAGCAGATAGGTAGCAAAATT 232
QY 30 LysGlyValPheLysSerThrMetGluGlyAlaMetLeuProAsnTyrGlyValala 49
Db 233 AAAGGAGTATTCAAAAGTACCACATGGAGGAGCTATGTCCTAAATTATGGTGTAGCT 292
QY 50 GluGlyGluAspPheIleGluGlyIleValValMetGluAspSerProVal 69
Db 293 GAAGGTGAAGATCATTTTATGAAGAAGTATTTGTTGTAATGGAAGATGATTCTCCAGTG 352
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaAlaIlePheIlePro 89
Db 353 GAGGCTGTGAGCACACCTAATATCTCCCGAAACCTTCTGTCATGGAATAGCATTC 412
QY 90 TyrValAspPheGluAspProSerSerGluArgGlyGluLysGluArgIlePro 109
Db 413 TATGTAGACTTTTGTGAGGATCCTCTCTGGAAGGAGGAGAGAAAGAAAGATTCT 472
QY 110 ValPheCysIleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHis 129
Db 473 GTGTTTTGTATGATGTTGAAGAAATGATAGAGAGCAGTTGGACACGAGCTGAACAT 532
QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149
Db 533 TGGTCTCTCTAGAGATATCTTGAATCTATGATCTTGAATCAAACTTACAGAAATT 592
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169
Db 593 CATGGTGCATTTCTGTATGCCACTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 652
QY 170 GluPheLeuLysSerLysArgGluGluPheGluGluTyrLeuGlnLysLeuGlnHis 189
Db 653 GAATTCCTTAAGTCAAGAGGAGAGAGTTCCAGAAATATCTACAGAACTTCTGCAGCAT 712
QY 190 ProGluLeuSerAsnSerGlnLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209
Db 713 CCAGAACTGAGTAAATAGTCAACTCTCTGCGAGACTTCTTCTTCTTCTTCTTCTTCT 772
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValPro 229
Db 773 CAATTTCTGTAGTAAAGTAACTACCAAGATGTAATCTTGGAAATATTAATATCTGTTCT 832
QY 230 GlyLysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsn 249
Db 833 GGAATACTTAATGAAGAGAGAGGTTCAGATTTGGACCTTTTATCATGAAATTTCAATAA 892
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269
Db 893 TCTTGTGAGTCTCCAAAGCCTAAACCAAGTAGACAGCACTGACCAATCTCAGCCCTACT 952
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289
Db 953 TCAGAAAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATATGCAACCGTGCTGAA 1012

```

```

RESULT 2
US-10-037-270-922
; Sequence 922, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Trillinghaast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-03-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 922
; LENGTH: 1716
; TYPE: DNA

```

```

290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309
1013 AATACAGAGAGAGAAACAAAATCAGAAATTATTTATGGAGGTGATGATCTGTAGAGAGATC 1072
310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHis 329
1073 TATGATTACCTGATGATGATAGAGCGGTAGTTTCCAGGTTCCTGACTGGCTTCATCAT 1132
330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349
1133 CTCTTAATGGAACTCGAATCTCTTTAAACACACCTCGAAATGTATATGATTACTAT 1192
350 LeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeu 369
1193 CTTCACTGTAAACTAGAACAGCTATTTCAGGAGCACCGCTTGTCTCTCATCTAATACACT 1252
370 LeuArgAspAlaIlePheCysGlnAsnThrGluProArgSerLeuGlnAspLysGlnLys 389
1253 CTGAGAGATGCTATATCTCTGTGAACACTGAACTCGCTCTCTCCAGATTAAGCAAAA 1312
390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409
1313 GGAGCAAAACAGACTTTTGAAGAAATGATGATTTACATTCAGATCTGTTAGTCAAGTGT 1372
410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429
1373 ATGGTCAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTCATGGCTTACAGCAACA 1432
430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449
1433 GTACTCAACAGCAGCTGACTTATGTTTATTTGGAATGATGATGATCAGGAACTGTTTCCA 1492
450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTrpMet 465
1493 GAGTCAATTAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 1540

```











Db 545 ATTACTCTGTTAGAGATGCTATATTTTGTGAACCGCTGAACACGATCATTACATCTG 604  
Qy 387 LysGlnLysGlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeu 406  
Db 605 AAGCAGCAAGGCAAGCTTACATTTGAAGAAATGATCGCTATATTCAGATTGATT 664  
Qy 407 ValLysCysIleGlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeu 426  
Db 665 GGTAAATGATTTGCTGATGAAAGCTAAATATGAAGCATTCGACTTCTGTTTGGTGATG 724

RESULT 7

US-09-918-995-5183  
; Sequence 5183, Application US/09918995  
; Publication No. US20030073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; PRIOR FILING DATE: 1999-01-20  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 5183  
; LENGTH: 397  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-918-995-5183

Alignment Scores:  
Pred. No.: 1,67e-73 Length: 397  
Score: 687.00 Matches: 131  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 28.19% Indels: 0  
DB: 11 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-918-995-5183 (1-397)

Qy 311 AspTyrLeuMetTyrValGlyArgValPheGlnValProAspTrpLeuHisLeu 330  
Db 3 GATTACCTGATGATGATGAGCGGTAGTTTCCAGGTTCCCTGACCTGCTCATCATCTC 62  
Qy 331 LeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyrLeu 350  
Db 63 TTAATGGGAACCTGAAATCCCTCTTTAAACACCCCTGGAATGATATCTGATTACTATCT 122  
Qy 351 GlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeuIleThrLeuLeu 370  
Db 123 CAGTGTAACCTAGACAGCTATTTCCAGGACCGGTTTGTCTCTCATACACTTCTC 182  
Qy 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390  
Db 183 AGAGATGCTATATTTCTGTGAAACACCTGAACTCGCTCTCTCCAGATAAGCAAAAGGA 242  
Qy 391 AlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCysIle 410  
Db 243 GCATAACAGACTTTTGAAGAAATGATGAATATACATTCAGATCTGTTAGTCAAGTGAT 302  
Qy 411 GlyGluGlnThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnProVal 430  
Db 303 GGTGAAGAACCAAGATGAAGCATCAGACTTCTGTTTGTATGCTTACAGCAACCGATGA 362  
Qy 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAsp 441  
Db 363 CTCACACAGGAGCTGACTTATGTTTATTGGAC 395

RESULT 8

US-09-983-965-540  
; Sequence 540, Application US/09983965  
; Patent No. US20020137160A1

; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Mengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR APPLICATION NUMBER: US 09/465,231  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR APPLICATION NUMBER: US 60/113,678  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 540  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURES:  
; NAME/KEY: unsure  
; LOCATION: (339)  
; OTHER INFORMATION:  
; OTHER INFORMATION: Clone ID: 64-BOVMS1-017-Q1-E1-H12  
US-09-983-965-540

Alignment Scores:  
Pred. No.: 6.9e-73 Length: 402  
Score: 682.00 Matches: 127  
Percent Similarity: 99.25% Conservative: 5  
Best Local Similarity: 95.49% Mismatches: 1  
Query Match: 27.99% Indels: 0  
DB: 10 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-983-965-540 (1-402)

Qy 205 AsnGlyGlyGluThrGlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIle 224  
Db 3 AATGTTGGGAAACACACAGTTTCTTTGATAGATACCTACGACGCTAAATCTTTGGAAATT 62  
Qy 225 IleLysSerValProGlyLysLeuMetLysGlyGlnHisLeuGluProPheIle 244  
Db 63 ATAAATCTGTTCTTGAAACTTAATGAAGAGAAAGTCAACATTTGGAACTTTCATC 122  
Qy 245 MetAsnPheIleAsnSerCysGluSerProLysProLysProSerArgProGluLeuThr 264  
Db 123 ATGATTTTCAATTAATCTTTGTAATCTCCAAAGCTAAACCGAGTAAACCGAACTGACC 182  
Qy 265 IleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsn 284  
Db 183 ATTTCTAGCCCTACCTCAGAGAAATAATAAGAAAGCTTTTAAATGATCTGTATAAGAAATA 242  
Qy 285 AlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMet 304  
Db 243 GCRAACCGTCTGAGATACAGAAAGAGGCAAAATCAGATTAATTTTATGGAATGATG 302  
Qy 305 ThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValValPheGlnValPro 324  
Db 303 ACTGTAGAGAGGCTTATGATTACTGATGTATGTATGTATGTATGTATGTATGTATGTAT 362  
Qy 325 AspTrpLeuHisLeuLeuMetGlyThrArgIleLeu 337  
Db 363 GACTGGCTTCATCATCTCTTAATGGAACTCGAATCCTC 401

RESULT 9

US-10-242-535A-25415  
; Sequence 25415, Application US/10242535A  
; Publication No. US20040013663A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.  
; APPLICANT: Liaw, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2005

;; CURRENT APPLICATION NUMBER: US/10/242,535A  
;; CURRENT FILING DATE: 2002-09-12  
;; PRIOR APPLICATION NUMBER: US 10/085,783  
;; PRIOR FILING DATE: 2002-02-28  
;; PRIOR APPLICATION NUMBER: US 60/305,340  
;; PRIOR FILING DATE: 2001-07-13  
;; PRIOR APPLICATION NUMBER: US 60/275,017  
;; PRIOR FILING DATE: 2001-03-12  
;; PRIOR APPLICATION NUMBER: US 60/271,955  
;; PRIOR FILING DATE: 2001-02-28  
;; NUMBER OF SEQ ID NOS: 58994  
;; SOFTWARE: PatentIn version 3.2  
;; SEQ ID NO 25415  
;; LENGTH: 471  
;; TYPE: DNA  
;; ORGANISM: Human  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (17)..(17)  
;; OTHER INFORMATION: n is a, c, g, or t  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (19)..(19)  
;; OTHER INFORMATION: n is a, c, g, or t  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (33)..(33)  
;; OTHER INFORMATION: n is a, c, g, or t  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (49)..(49)  
;; OTHER INFORMATION: n is a, c, g, or t  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (58)..(59)  
;; OTHER INFORMATION: n is a, c, g, or t

## US-10-242-535A-25415

Alignment Scores:  
Pred. No.: 5,15e-69 Length: 471  
Score: 651.00 Matches: 131  
Percent Similarity: 94.24% Conservatives: 0  
Best Local Similarity: 94.24% Mismatches: 7  
Query Match: 26.71% Indels: 1  
DB: 12 Gaps: 0

US-09-744-313A-1 (1-465) x US-10-242-535A-25415 (1-471)

QY 328 HisHisLeuLeuMetGlyThrArgIleLeuPhelysAsnThrLeuGluMetTyr-ThrAs 347  
DB 4 CATGACCTGTAAANGNGAACTCGAATCCTNTTTAAAAACACCCCTGNAATGTATNNTGA 63  
QY 347 pTyrTyrLeuGlnCysIysLeuGlnLeuPheGlnGluHisArgLeuValSerLeu11 367  
DB 64 TTACTATCTTCAGTGTAAACAGACAGCTATTTTCAAGAGACCCGTTTGTCTCATCTCAT 123  
QY 367 eThrLeuLeuArgaspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLy 387  
DB 124 AACACTTCTCAGAGATGCTATATCTGTGAAACACCTGACCTCGCTCTCCAGATAA 183  
QY 387 sGlnIysGlyAlaIysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuVal 407  
DB 184 GCMAAAGGAGCAAAACAGACTTTTGAAGAAATGATGATATACATTCAGATCTGTAGT 243  
QY 407 lIysCysIleGlyGluGluThrIysTyrGluSerIleArgLeuLeuPheAspGlyLeuGI 427  
DB 244 CAAGTGTATTGGTGAAGAAACCAAGTATGAAAGCATCAGACCTCTGTTTGTGGCTTACA 303  
QY 427 nglnProValLeuAsnIysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLe 447  
DB 304 GCACCACTACTCAACAGCAGCTGACTTATGTTTATTTGACATTTGTGATCAGGAACT 363  
QY 447 uPheProGluLeuAsnIysValGlnIysGluValThrSerValThrSerTrpMet 465

DB 364 GTTTCAGAGCTCAATAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 418

## RESULT 10

US-09-918-995-20205  
;; Sequence 355, Application US/09918995  
;; Publication No. US20030073623A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Hyseq, Inc.  
;; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
;; FILE REFERENCE: 20411-756  
;; CURRENT APPLICATION NUMBER: US/09/918,995  
;; CURRENT FILING DATE: 2001-07-30  
;; PRIOR APPLICATION NUMBER: US/09/235,076  
;; PRIOR FILING DATE: 1999-01-20  
;; NUMBER OF SEQ ID NOS: 38054  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 20205  
;; LENGTH: 508  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-918-995-20205

Alignment Scores:  
Pred. No.: 5,5e-56 Length: 508  
Score: 544.00 Matches: 107  
Percent Similarity: 100.00% Conservatives: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 22.32% Indels: 0  
DB: 11 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-918-995-20205 (1-508)

QY 359 GlnGluHisArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePheCysGluAsn 378  
DB 65 CAGGAGCACCGCTTGGTCTCACTCAATACACTTCTCAGAGATGCTATATTCTGTGAAAC 124  
QY 379 ThrGluProArgSerLeuGlnAspIysGlnIysGlyAlaIysGlnThrPheGluGluMet 398  
DB 125 ACTGAACCTCGCTCTCTCCAGATAAGCAAAAGGAGCAAAACAGACTTTTGAAGAATG 184  
QY 399 MetAsnTyrIleProAspLeuValIysCysIleGlyGluGluThrLysTyrGluSer 418  
DB 185 ATGAATTACATTCAGATCTCTTAGTCAAGTGTATTGTTGAAGAACCAAGTATGAAGC 244  
QY 419 IleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnIysGlnLeuThrTyrVal 438  
DB 245 ATCAGACTTCTGTTGTATGGCTTACAGCAACCACTACTCAACAGCAGCTGACTTATGTT 304  
QY 439 LeuLeuAspIleValIleGlnGluLeuPheProGluLeuAsnIysValGlnIysGluVal 458  
DB 305 TTATTGGACATTTGTGATACAGGAACCTTTTCCAGAGCTCAATAAGGTACAAAGGAAGTT 364  
QY 459 ThrSerValThrSerTrpMet 465  
DB 365 ACCTCTGTGACATCTTGGATG 385

## RESULT 11

US-09-983-965-355  
;; Sequence 355, Application US/09983965  
;; Patent No. US20020137160A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Warren, Wesley C.  
;; APPLICANT: Tao, Nengbing  
;; APPLICANT: Byatt, John C.  
;; APPLICANT: Mathialagan, Nagappan  
;; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
;; FILE REFERENCE: 37-21(10297)C  
;; CURRENT APPLICATION NUMBER: US/09/983,965  
;; CURRENT FILING DATE: 2001-10-26  
;; PRIOR APPLICATION NUMBER: US 09/465,231

```
; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 355
; LENGTH: 305
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 42-BOVMS1-019-Q1-E1-C6
US-09-983-965-355

Alignment Scores:
Pred. No.: 1,26e-51 Length: 305
Score: 505.00 Matches: 97
Percent Similarity: 97.03% Conservative: 1
Best Local Similarity: 96.04% Mismatches: 3
Query Match: 20.72% Indels: 0
DB: 10 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-983-965-355 (1-305)
QY 180 GlnGluTyrLeuGlnLysLeuLeuGlnHisProGluLeuSerAsnSerGlnLeuLeuAla 199
DB 1 CAGGAATATCTGCGAAGAACTTCTGCAACATCCAGATTGATGATACAGTCACTTCTGGCA 60

QY 200 AspPheLeuSerProAsnGlyGlyGluThrGlnPheLeuAspLysLeuProAspVal 219
DB 61 GATTTCTTCTCCCAATGTTGGGGAACACAGTTTCTTGATTAAGATATACACAGCTA 120

QY 220 AsnLeuGlyLysLleLleLysSerValProGlyLysLeuMetLysGlyGlnHis 239
DB 121 AATCTTGGGAAAATATATAATCTTCTCGGAAAACATAATGAAGAGAAAGGTCAACAT 180

QY 240 LeuGluProPheLleMetAsnPheLleAsnSerCysGluSerProLysProLysProSer 259
DB 181 TTGAACCTTTCATCATGAATTTCTTAATCTTGTGAATCTCAAGCCTTAACCGAGT 240

QY 260 ArgProGluLeuThrLleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAsp 279
DB 241 AAACAGAACTGACCAATCTCAGCCCTACATCAGAGATATATAACAGCTTTTATGAT 300

QY 280 Leu 280
DB 301 CTG 303

RESULT 12
US-10-242-535A-49594
; Sequence 49594, Application US/10242535A
; Publication No. US20040013663A1
; GENERAL INFORMATION:
; APPLICANT: ChondroGene Inc.
; APPLICANT: Liaw, C.C.
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis
; FILE REFERENCE: 4231/2005
; CURRENT APPLICATION NUMBER: US/10/242,535A
; PRIOR FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 10/085,783
; PRIOR FILING DATE: 2002-02-28
; PRIOR APPLICATION NUMBER: US 60/305,340
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/275,017
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: US 60/271,955
; PRIOR FILING DATE: 2001-02-28
; NUMBER OF SEQ ID NOS: 58994
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 49594
; LENGTH: 446
; TYPE: DNA
; ORGANISM: Human
US-10-242-535A-49594

; PRIOR FILING DATE: 1999-12-15
; PRIOR APPLICATION NUMBER: US 60/113,678
; PRIOR FILING DATE: 1998-12-17
; NUMBER OF SEQ ID NOS: 5912
; SEQ ID NO 355
; LENGTH: 305
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; OTHER INFORMATION: Clone ID: 42-BOVMS1-019-Q1-E1-C6
US-09-983-965-355

Alignment Scores:
Pred. No.: 2.73e-49 Length: 446
Score: 488.00 Matches: 95
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 20.02% Indels: 0
DB: 12 Gaps: 0

US-09-744-313A-1 (1-465) x US-10-242-535A-49594 (1-446)
QY 371 ArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLysGly 390
DB 1 AGAGATGCTATATCTCTGTGAAACACACTGAACCTCGCTCTCTCCAGATAAGCAAAAGGA 60

QY 391 AlaLysGlnThrPheGluGluMetMetAsnTyrLleProAspLeuLeuValLysCysLle 410
DB 61 GCAAAACACAGACTTTTGAAGAAATGATGAATTACATTCAGATCTGTGTAGTCACTGATT 120

QY 411 GlyGluGluThrLysTyrGluSerLleArgLeuLeuPheAspGlyLeuGlnProVal 430
DB 121 GGTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTGTGATGCTTACAGCAACCAAGTA 180

QY 431 LeuAsnLysGlnLeuThrTyrValLeuLeuAspLleValleGlnGluLeuPheProGlu 450
DB 181 CTCACACAGCAGCTGACTTATGTTTATTTGACATTTGTGATACAGGAACCTGTTCCAGAG 240

QY 451 LeuAsnLysValGlnLysGluValThrSerValThrSerTyrMet 465
DB 241 CTCATTAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 285

RESULT 13
US-10-108-260A-839
; Sequence 839, Application US/10108260A
; Publication No. US20040005560A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20040005560A1el full length cDNA
; FILE REFERENCE: H1-A0106
; CURRENT APPLICATION NUMBER: US/10/108,260A
; CURRENT FILING DATE: 2002-03-27
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 839
; LENGTH: 2720
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-108-260A-839

Alignment Scores:
Pred. No.: 1.41e-20 Length: 2720
Score: 262.00 Matches: 98
Percent Similarity: 42.65% Conservative: 79
Best Local Similarity: 23.61% Mismatches: 154
Query Match: 10.75% Indels: 84
DB: 12 Gaps: 15

US-09-744-313A-1 (1-465) x US-10-108-260A-839 (1-2720)
QY 56 lleGluGluGlyIleValValMetGlu-----AspAspSerProValGluAlaVal 72
DB 1133 TTGAAGGATGAATAATCTTAATAGAGAAGACGACACCTTCAGCTGCACATGGCA 1192

QY 73 SerThrProAsnThrProArgAsnLeuAlaAlaTyrLysLleSerIleProTyr-ValAsp 92
DB 1193 AGAACGGATTGGTGTGTGAAACCTTGGCATGTGGAAAGCCTCAUCCAGTGGAGAG 1252

QY 93 PhePheGluAspProSerSerGluArgLysGluLysGluArgLleProValPheCys 112
DB 1253 GTTACAGAA-----GAGATGGTGAGCAATTGCCATGTTACTTT 1291

QY 113 lleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisTyrSerVal 132
DB 1292 GTCATGGTAAAGCCTACAAGAA-----GTTGGAGGAGTTGAAACTAGAACTGGACGTC 1345
```

QY 133 TyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAla 152  
Db 1346 CCAGAGGCTCAGCAGATTTCAGATTTCACCGGAAACTCAGTGAGTGCCTCTCT 1405  
QY 153 PheProAspAlaGlnLeuProSerLysArgIleGlyProLysAsnTyrGlu----- 170  
Db 1406 TTAATAAAAGTCAGTTCCTCTCTTAGCAAGCTGCTTTCATAATCTATAGATCAAAAG 1465  
QY 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190  
Db 1466 TTTATGAAAGTCGAGATCAATTAATAAGTTTTTACAGAACTGCTTTTCAGATGAA 1525  
QY 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210  
Db 1526 AGACTGTGTCAGAGTGAAGCTCTTTATGCTTCTTGAAGCTTCTCTGACTACCTCAAG 1585  
QY 211 PheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLeuLysSerValProGly 230  
Db 1586 GTTATCGAC-----GTGCGAGGG 1603  
QY 231 LysLeuMetLysGlnLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSer 250  
Db 1604 -----AAAAAAATCTTTTTCATTATCTCTGTTTTTG----- 1636  
QY 251 CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer 270  
Db 1636 ----- 1636  
QY 271 GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsn 290  
Db 1637 -----GAAGACTTCTCGGACTTCTTC-----TCCACCCAGGAGGAGG 1678  
QY 291 ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr 310  
Db 1679 ACAGAGGAGGACAGTGCCTGTCAGATTATGTTGAT-----GATGTGGATGGAGGAAA 1732  
QY 311 AspTyrLeu-----MetTyrValGlyArgValPheGlnValPro--- 324  
Db 1733 GACGCTTGGCTGAACCATGTTTCATGTTGATGGGAG---ATTITGAACCTCGAGGA 1789  
QY 325 -----AspTyrLeuHisLeuLeuMetGly---ThrArgIleLeuPheLysAsnThr 341  
Db 1790 ATGTTTAAATGGGTCAGAGACACTTAATGCTCTGCTCAGTGCATCTTTGGAGNACC 1849  
QY 342 LeuGluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHis 361  
Db 1850 ATC-----AACAAACAAATCGGACACAGTGCAGTGGATTTCAGTGAGCAA 1897  
QY 362 ArgLeuValSerLeuIleThrLeuLeuArgAspAlaIlePhe-----Cys 376  
Db 1898 ATGTTGGTTACTACATCAATATTTTCGGGATGCTTTTGGCCAAATGGAAAGTTGGCA 1957  
QY 377 GluAsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGlu 396  
Db 1958 CCACGACCAATCAGAAAGCAAGCAAGTGCAGAAACAAACAGAGACGACAGCAA 2017  
QY 397 GluMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyr 416  
Db 2018 AAGCTGCTGAAACCAATTCAGATATGCTTCAGAGCCTTGTGGACAGCAAAATGCCCGC 2077  
QY 417 GluSerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThr 436  
Db 2078 CACGGTATAATAAATAATTCATGCACTGCAAGAAACAGAGCCAAACAGCATCTCTTA 2137  
QY 437 TyrValLeuLeuAspIleValIleGlnGluLeuPheProGluLeu 451  
Db 2138 TATGCGCTGATGGAACTGCTGCTAATTAATGAACTGTGCTGAGCTG 2182

RESULT 14

US-09-822-846-238

; Sequence 238, Application US/09822846

; Publication No. US20030027139A1

GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: LaVallie, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steininger II, Robert J.  
; APPLICANT: Bowman, Michael R.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakar  
; APPLICANT: Graham, James R.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6400  
; CURRENT APPLICATION NUMBER: US/09/822,846  
; CURRENT FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195,605  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 629  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 238  
; LENGTH: 2494  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-846-238

Alignment Scores:  
Pred. No.: 4e-19 Length: 2494  
Score: 249.50 Matches: 100  
Percent Similarity: 42.75% Conservative: 77  
Best Local Similarity: 24.15% Mismatches: 155  
Query Match: 10.24% Indels: 83  
DB: 11 Gaps: 14

US-09-744-313A-1 (1-465) x US-09-822-846-238 (1-2494)

QY 56 IleGluGluGlyIleValValMetGlu-----AspAspSerProValGluAlaVal 72  
Db 908 TTGAAGATGAATAATCTTAATAGAGAAAGACGACAGCTTCAGTCCATGGCA 967  
QY 73 SerThrProAsnThrProArgAsnLeuAlaAlaIleLysIleSerIleProTyrValAsp 92  
Db 968 AGAACGGATTGGTGTGTGAAACCTTGGCATGTGGAAAGCCTCCATCCAGTGGAGAG 1027  
QY 93 PhePheGluAspProSerSerGluArgLysGluLysLysGluArgIleProValPheCys 112  
Db 1028 GTTACAGAA-----GAGAATGGTGGAGCAATTGCCATGTACTTT 1066  
QY 113 IleAspValGluArgAsnAspArgAlaValGlyHisGluProGluHisIleTyrSerVal 132  
Db 1067 GTCATGGTAAGCCTACAAGAA-----GTTGGAGAGTGTGAACCTAAGACTGGACGTC 1120  
QY 133 TyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPheHisGlyAla 152  
Db 1121 CCCAGAGGCTCAGCAGAGTTTCAGAAATTTACCGGAAACTCAGTGAAGTGCCTCTCT 1180  
QY 153 PheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyrGlu----- 170  
Db 1181 TTAATAAAAGTCCAGTTGCTTCTCTTAGCAAGCTGCCTTTCAAATCTATAGATCAAAAG 1240  
QY 171 PheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHisPro 190  
Db 1241 TTTATGAAAGTCCGAGAACTCAATTAATAAGTTTTTACAGAACTCTCTTCAGATGAA 1300  
QY 191 GluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThrGln 210

Db	1301	AGACTGCTCAGAGTGAAGC	TTTATGCTCTTGTAGCCCTTCTCTGACTACCTCAAG	1360
Qy	211	PhelLeuAspLysIleLeuProAspValAsnLeuGlyLysIleIleLysSerValProGly	230	
Db	1361	GTATCGAC	-----GTCCAGGGG	1378
Qy	231	LysLeuMetLysGluLysGlyGlnHisLeuGluProPheIleMetAsnPheIleAsnSer	250	
Db	1379	-----AAAAAATTCCTTTTCATTATTCCTCATTG	-----1411	
Qy	251	CysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThrSer	270	
Db	1411	-----	-----1411	
Qy	271	GluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAlaAsnArgAlaGluAsn	290	
Db	1412	-----GAAAGACTTCTCGCACTTCTTC	-----TCCACACAGGAGGAGGAG 1453	
Qy	291	ThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyValTyr	310	
Db	1454	ACAGAGGAGCAGTGACCTGTCCAGATTATGTCAT	-----GATGTGGATGGAGGAAA 1507	
Qy	311	AspTyrLeu	-----MetTyrValGlyArgValPheGlnValPro--- 324	
Db	1508	GAGCGCTTGGCTGAACCATGTTTCATGTTGATGGGGAG	-----ATTTTGAACCTTCGAGGA 1564	
Qy	325	-----AspTrpLeuHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeu	342	
Db	1565	ATGTTTAAATGGGTGAGAGAACATTAATTGCC	-----CTCGTTCAGGTCACCTTG 1615	
Qy	343	GluMetTyrThrAspTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArg	362	
Db	1616	GAAGA-ACCATCAACAACAATCCGGGACACAGTCAGTCAGTGGATTTTCATGAGCAATG	1674	
Qy	363	LeuValSerLeuIleThrLeuLeuArgAspAlaIlePhe	-----CysGlu 377	
Db	1675	TTGTTTACTACATCAATATTTTCCGGCATGCTTTTGGCCAAATGGGAAGTTGGCACCA	1734	
Qy	378	AsnThrGluProArgSerLeuGlnAspLysGlnLysGlyAlaLysGlnThrPheGluGlu	397	
Db	1735	CCGACCAACAATCAGAAGCAAGAGCAAGTCAGGAAAA	-----CAAAACAGAGACACAGCAAAAG 1794	
Qy	398	MetMetAsnTyrIleProAspLeuLeuValLysCysIleGlyGluGluThrLysTyrGlu	417	
Db	1795	CTGCTTGAAACATTCAGATATGCTTCAGAGCCCTTGTGGCAGCAAAATGCCCGCCAC	1854	
Qy	418	SerIleArgLeuLeuPheAspGlyLeuGlnGlnProValLeuAsnLysGlnLeuThrTyr	437	
Db	1855	GGTATATAAATAATTCAATGCACTGCAGAAACAGAGCCACACAGCATCTGTTATAT	1914	
Qy	438	ValLeuLeuAspIleValIleGlnGluLeuPheProGluLeu	451	
Db	1915	GCCTGATGAACACTGCTGCTAATGAACCTGTGCTCTGAGTGTG	1956	

RESULT 15

US-09-783-590-4866  
; Sequence 4866, Application US/09783590  
; Patent No. US20020110850A1  
; GENERAL INFORMATION:  
; APPLICANT: Dillan, Patrick J.  
; APPLICANT: Haseltine, William A.  
; APPLICANT: Li, Haodong  
; APPLICANT: Rosen, Craig A.  
; APPLICANT: Ruben, Steven M.  
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
; FILE REFERENCE: PO-16.2C1  
; CURRENT APPLICATION NUMBER: US/09/783,590  
; CURRENT FILING DATE: 2000-02-15  
; PRIOR APPLICATION NUMBER: 08/420,856  
; PRIOR FILING DATE: 1995-04-12  
; PRIOR APPLICATION NUMBER: 08/346,731  
; PRIOR FILING DATE: 1994-11-21

```

; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4866
; LENGTH: 280
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (17)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (50)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (92)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (127)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (146)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (171)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (188)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (208)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (220)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (233)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (253)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (274)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (275)
; OTHER INFORMATION: n equals a,t,g, or c
; US-09-783-590-4966

```

Alignment Scores:	
Pred. No.:	1,586-20
Score:	248.00
Percent Similarity:	77.91%
Best Local Similarity:	73.26%
Query Match:	10.18%
DB:	10
	2
Length:	280
Matches:	63
Conservative:	4
Mismatches:	12
Indels:	7
Gaps:	2

US-09-744-313A-1 (1-465) x US-09-783-590-4866 (1-280)

Qy	264	ThrIleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsn-AspLeuPheLysAs	283
Db	3	ACCAATCTCAGCCNACTTCAGAAAACAACAAGAAGCTTTTCAATGANTCTGTTTAAAAA	62
Qy	283	AsnAla-AsnArgAlaGluAsnThrGluArg--LysGlnAsnGlnAsnTy-PheMetG	302
Db	63	TAATGCCAACCGTCTCGAAATAATCAGAGNGATAAACAACATCAGAAATATTTTATGGA	122
Qy	302	uValMetThrValGluGlyValTyAspTyTyr---LeuMetTyTyrValGlyArgValValPh	321
Db	123	GGTGNTGACTGTAGAAGCAGGTGTGTGAATTACCTGAATGTGTAGGCGGGTAGTGT	182
Qy	321	eGlnValProAspTyrPheHisLeuLeuMetGlyThrArgIleLeu----PheLysA	340
Db	183	CAGAGNCTCGACTGGCTTCATCTATCTTTAATGGGNGCTCAGTAATCNCCTTTTAAAA	242

Oy 340 snThrieu 342  
| | | | |  
Db 243 ACACCTG 250

Search completed: January 31, 2004, 06:31:34  
Job time : 523 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 15:56:14 ; Search time 125 Seconds  
(without alignments)  
7033.882 Million cell updates/sec

Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: 1-gtataaaactccaaaagt.....tttaataataaaaaaaaaa 1992

Scoring table: OLIGO\_NUC  
Gapop 60.0  
Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Issued Patents NA.\*  
1: /cgn2\_6/prodata/2/ina/5A-COMB.seq.\*  
2: /cgn2\_6/prodata/2/ina/5B-COMB.seq.\*  
3: /cgn2\_6/prodata/2/ina/6A-COMB.seq.\*  
4: /cgn2\_6/prodata/2/ina/6B-COMB.seq.\*  
5: /cgn2\_6/prodata/2/ina/PCTUS-COMB.seq.\*  
6: /cgn2\_6/prodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1532	76.9	1716	4	US-09-620-312D-922
2	918	46.1	1551	4	US-09-620-312D-923
3	20	1.0	271	2	US-08-731-272A-29
4	20	1.0	7015	3	US-09-177-249-6
5	19	1.0	731	1	US-08-451-405A-2
6	19	1.0	2060	1	US-07-721-761A-31
7	19	1.0	2060	1	US-07-978-687-31
8	19	1.0	2060	5	PCT-US91-01750-2
9	19	1.0	2060	5	PCT-US91-05801-31
10	19	1.0	2081	5	PCT-US91-01750-3
11	19	1.0	3440	1	US-08-471-791-27
12	19	1.0	3440	5	PCT-US91-01746-27
13	19	1.0	3826	4	US-09-220-132-51
14	18	0.9	141	3	US-08-737-078A-1
15	18	0.9	141	5	PCT-US94-04705-1
16	18	0.9	243	4	US-09-134-001C-671
17	18	0.9	544	4	US-09-280-116-247
18	18	0.9	632	4	US-09-489-847-56
19	18	0.9	1001	4	US-09-841-638-339
20	18	0.9	1001	4	US-09-641-638-363
21	18	0.9	1001	4	US-09-641-638-600
22	18	0.9	1395	4	US-09-134-001C-651
23	18	0.9	1926	4	US-09-613-303-50
24	18	0.9	2356	1	US-08-105-483-222
25	18	0.9	2356	1	US-08-220-151-75
26	18	0.9	2356	1	US-08-413-118-75
27	18	0.9	2356	1	US-08-224-657-51

C 28	18	0.9	2356	1	US-08-709-209-222	Sequence 222, App
C 29	18	0.9	2356	1	US-08-458-101-222	Sequence 222, App
C 30	18	0.9	2356	2	US-08-184-009-78	Sequence 78, Appl
C 31	18	0.9	2356	2	US-08-417-210A-68	Sequence 68, Appl
C 32	18	0.9	2356	2	US-08-458-356-78	Sequence 78, Appl
C 33	18	0.9	2356	3	US-08-473-446-75	Sequence 75, Appl
C 34	18	0.9	2356	3	US-08-460-736-78	Sequence 78, Appl
C 35	18	0.9	2356	4	US-09-354-138-51	Sequence 51, Appl
C 36	18	0.9	2356	4	US-09-535-370-78	Sequence 78, Appl
C 37	18	0.9	2760	1	US-08-101-593-1	Sequence 1, Appl
C 38	18	0.9	2760	1	US-08-101-593-3	Sequence 3, Appl
C 39	18	0.9	3475	4	US-08-924-629C-15	Sequence 15, Appl
C 40	18	0.9	4072	3	US-09-272-496-7	Sequence 7, Appl
C 41	18	0.9	4206	4	US-09-302-620B-81	Sequence 81, Appl
C 42	18	0.9	4206	4	US-09-312-161-3	Sequence 3, Appl
C 43	18	0.9	5385	4	US-08-361-527-77	Sequence 77, Appl
C 44	18	0.9	5796	4	US-09-366-715-4	Sequence 4, Appl
C 45	18	0.9	10614	1	US-08-135-511-35	Sequence 35, Appl

ALIGNMENTS

RESULT 1  
US-09-620-312D-922  
; Sequence 922, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunging  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghaast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes Version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (71)..(1543)  
US-09-620-312D-922

Query Match 76.9%; Score 1532; DB 4; Length 1716;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 147 GGACACACAGAAAGGGGAGATCATTTGGATCAGACAGATAGGTAGCAATTAAG 206  
Db 177 GGACACACAGAAAGGGGAGATCATTTGGATCAGACAGATAGGTAGCAATTAAG 236

207 GAGTATTTCAAAAGTACCAACATGGAGGAGCTATGTTGCTTAATTAATGTTGAGCTGAAG 266  
237 GAGTATTTCAAAAGTACCAACATGGAGGAGCTATGTTGCTTAATTAATGTTGAGCTGAAG 296  
267 GTGAAGATGATTTTATTTAGGAAGGTATTTGTTGTAATGGAAGATGATTTCTCCAGTGGAGG 326  
297 GTGAAGATGATTTTATTTAGGAAGGTATTTGTTGTAATGGAAGATGATTTCTCCAGTGGAGG 356  
327 CTGTGAGCAACCTAATATCTCCCGAAACCTTGCTGTCATGGAATAATAGCATTCATATG 386  
357 CTGTGAGCAACCTAATATCTCCCGAAACCTTGCTGTCATGGAATAATAGCATTCATATG 416  
387 TAGACTTTTTTGGAGTCCCTCTCTGGAAGGAGGAGGAAGGAAGAAATTCCTGTGT 446  
417 TAGACTTTTTTGGAGTCCCTCTCTGGAAGGAGGAGGAAGGAAGAAATTCCTGTGT 476  
447 TTTGTAATGATGTTGGAAGAAATGATGGAAGAGCAGTTCGACACAGAGCCTGAAACATTCGT 506  
477 TTTGTAATGATGTTGGAAGAAATGATGGAAGAGCAGTTCGACACAGAGCCTGAAACATTCGT 536  
507 CTGCTATGAGAGATATCTTGAAATTCATGTAATCAAACTTAACAGAAATTCATG 566  
537 CTGCTATGAGAGATATCTTGAAATTCATGTAATCAAACTTAACAGAAATTCATG 596  
567 GTGCATTTCTGATGCCAGCTTCCTCTTAAGAGGATCATTCGCCCCCAAAATTAATGAT 626  
597 GTGCATTTCTGATGCCAGCTTCCTCTTAAGAGGATCATTCGCCCCCAAAATTAATGAT 656  
627 TCTTAAGTCAAGAGGAGAGGTCCAGAGATATCTACAGAAATTCCTGACGATCCAG 686  
657 TCTTAAGTCAAGAGGAGAGGTCCAGAGATATCTACAGAAATTCCTGACGATCCAG 716  
687 AACTGAGTAATAGTCAACTTTCTGGCAGACTTTCTTCCCTTAATGTTGGGGGAAACACAAAT 746  
717 AACTGAGTAATAGTCAACTTTCTGGCAGACTTTCTTCCCTTAATGTTGGGGGAAACACAAAT 776  
747 TCTTGAATGAGATACCTACGAGATGTAATCTTGGGAAATTAATAATCTGTTCTCGAA 806  
777 TCTTGAATGAGATACCTACGAGATGTAATCTTGGGAAATTAATAATCTGTTCTCGAA 836  
807 AACTGATGAAGAGAGAGTTCAGATTTGGAACCTTTTATCATGATTTTCAATATTCCT 866  
837 AACTGATGAAGAGAGAGTTCAGATTTGGAACCTTTTATCATGATTTTCAATATTCCT 896  
867 GTGAGTCTCAAAAGCTTTTCAATGATCTGTTTAAATAATGCAACCCGCTGCAAAATA 926  
897 GTGAGTCTCAAAAGCTTTTCAATGATCTGTTTAAATAATGCAACCCGCTGCAAAATA 956  
927 AAAACAAAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCCGCTGCAAAATA 986  
957 AAAACAAAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCCGCTGCAAAATA 1016  
987 CAGAGAGAGCAAAATCAGAAATTTATGAGAGTGTGATCTGTAAGAGAGTCTATG 1046  
1017 CAGAGAGAGCAAAATCAGAAATTTATGAGAGTGTGATCTGTAAGAGAGTCTATG 1076  
1047 ATTACCTGATGATGTAAGAGAGTGTGATTTTCAAGGTTCTGACTGGTTCATCTCTCT 1106  
1077 ATTACCTGATGATGTAAGAGAGTGTGATTTTCAAGGTTCTGACTGGTTCATCTCTCT 1136  
1107 TAAATGGGAACCTCAATCTCTTTTAAACACCCCTGGAATGTATGATCTATCTCTTC 1166  
1137 TAAATGGGAACCTCAATCTCTTTTAAACACCCCTGGAATGTATGATCTATCTCTTC 1196  
1167 AGTGTAACTAGAACAGCTATTTTCAAGAGCACCCTGTTTGTCTCAGTCATCACTCTCTCA 1226  
1197 AGTGTAACTAGAACAGCTATTTTCAAGAGCACCCTGTTTGTCTCAGTCATCACTCTCTCA 1256  
1227 GAGATGCTATATCTGTGGAACACCTCAACCTGCTCTCTCCAGATTAAGCAAAAGGAG 1286  
1257 GAGATGCTATATCTGTGGAACACCTCAACCTGCTCTCTCCAGATTAAGCAAAAGGAG 1316  
1287 CAAAACAGAGCTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGTATG 1346

1317 CAAAACAGAGCTTTTGAAGAAATGATGAATTAATTCAGATCTGTTAGTCAAGTGTATG 1376  
1347 GTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGAATGCTTACAGCAACCAAGTAC 1406  
1377 GTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTTGAATGCTTACAGCAACCAAGTAC 1436  
1407 TCAACAGAGCAGCTGACTTATGTTTATTTGACATCTGATACAGGAACCTGTTTCCAGAGC 1466  
1437 TCAACAGAGCAGCTGACTTATGTTTATTTGACATCTGATACAGGAACCTGTTTCCAGAGC 1496  
1467 TCAATAGGTACAAAAGAGGAGTTTACCTCTGTCGATCATCTTGGATGTAACACTTGGATTTG 1526  
1497 TCAATAGGTACAAAAGAGGAGTTTACCTCTGTCGATCATCTTGGATGTAACACTTGGATTTG 1556  
1527 GTATAGAAATACCAATTCGAAATTTCTGCTGTCGAGGTTGGTAGAAATTTACTTTTTTGG 1586  
1557 GTATAGAAATACCAATTCGAAATTTCTGCTGTCGAGGTTGGTAGAAATTTACTTTTTTGG 1616  
1587 GTATATTTCTTATATATATATGATGATCATCGCTGTCGAAATTTTAGTATTTTGTGTTTT 1646  
1617 GTATATTTCTTATATATATATGATGATCATCGCTGTCGAAATTTTAGTATTTTGTGTTTT 1676  
1647 AATAAGAGCTAACACAAACCTTAATGATTAATA 1678  
1677 AATAAGAGCTAACACAAACCTTAATGATTAATA 1708

## RESULT 2

US-09-620-312D-923  
; Sequence 923, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyang  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yuning  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP28  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL genes Version 1.0  
; SEQ ID NO 923  
; LENGTH: 1551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (71)..(1378)  
US-09-620-312D-923

Query Match 46.1%; Score 918; DB 4; Length 1551;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



```

TYPE: DNA
ORGANISM: Arabidopsis sp.
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(7014)
OTHER INFORMATION: fertilization-independent endospore 1 (FIE1)
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 1
FEATURE:
NAME/KEY: CDS
LOCATION: (2)..(7015)
OTHER INFORMATION: fertilization-independent endospore 1 (FIE1)
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 2
FEATURE:
NAME/KEY: CDS
LOCATION: (3)..(7013)
OTHER INFORMATION: fertilization-independent endospore 1 (FIE1)
OTHER INFORMATION: SET/polycomb gene genomic sequence reading frame 3
US-09-177-249-6
Query Match 1.0%; Score 20; DB 3; Length 7015;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 65 TTTTCTTTTAAATGTTTT 84
Db 5523 TTTTCTTTTAAATGTTTT 5542

RESULT 5
US-08-451-405A-2
Sequence 2, Application US/08451405A
Patent No. 5736358
GENERAL INFORMATION:
APPLICANT: FASEL, NICOLAS JOSEPH
APPLICANT: REYMOND, CHRISTOPHE DOMINIQUE
TITLE OF INVENTION: DICTYOSTELID EXPRESSION VECTOR AND
TITLE OF INVENTION: METHOD FOR EXPRESSING A DESIRED PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: THE WEBB LAW FIRM
CITY: PITTSBURGH
STATE: PENNSYLVANIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 15219-1818
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" FLOPPY DISK
COMPUTER: Microwest Micro 486-50
OPERATING SYSTEM: DOS
SOFTWARE: WORDPERFECT 6.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/451,405A
FILING DATE: 26-MAY-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 07/965,273
FILING DATE: 15-JAN-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 731
TYPE: NUCLEIC ACID
STRANDEDNESS: SINGLE
TOPOLOGY: UNKNOWN
US-08-451-405A-2
Query Match 1.0%; Score 19; DB 1; Length 731;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1307 TTTTCTTTTAAATGTTTT 1925
Db 71 TTTTCTTTTAAATGTTTT 89

us-09-721-761A-31
RESULT 6
US-07-721-761A-31
Sequence 31, Application US/07721761A
Patent No. 5475099
GENERAL INFORMATION:
APPLICANT: Vic. C. Knauf
APPLICANT: Gregory A. Thompson
TITLE OF INVENTION: Plant Fatty Acid Synthases
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Calgene, Inc.
STREET: 1920 Fifth Street
CITY: Davis
STATE: CA
COUNTRY: USA
ZIP: 95616
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.7
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/721,761A
FILING DATE: 19910626
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 07/568,493
FILING DATE: 15-AUGUST-1990
ATTORNEY/AGENT INFORMATION:
NAME: Elizabeth Laessen
REGISTRATION NUMBER: 31,845
NAME: Donna E. Scherer
REGISTRATION NUMBER: 34,719
REFERENCE/DOCKET NUMBER: CGNE 76-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (916) 753-6313
TELEFAX: (916) 753-1510
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 2060 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-07-721-761A-31
Query Match 1.0%; Score 19; DB 1; Length 2060;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1971 TTTTAAATAATAAAAAA 1989
Db 113 TTTTAAATAATAAAAAA 131

US-07-978-687-31
RESULT 7
US-07-978-687-31
Sequence 31, Application US/07978687
Patent No. 5510255
GENERAL INFORMATION:
APPLICANT: Vic. C. Knauf
APPLICANT: Gregory A. Thompson
TITLE OF INVENTION: Plant Fatty Acid Synthases
NUMBER OF SEQUENCES: 51
CORRESPONDENCE ADDRESS:
ADDRESSEE: Calgene, Inc.
STREET: 1920 Fifth Street
CITY: Davis
STATE: CA
COUNTRY: USA
ZIP: 95616
COMPUTER READABLE FORM:
```

MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.7  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
FILING DATE: FEBRUARY 1, 1993  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/05801  
FILING DATE: 15-AUGUST-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/568,493  
FILING DATE: 15-AUGUST-1990  
PRIOR APPLICATION DATA: 07/721,761  
FILING DATE: 26-JUNE-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Elizabeth Lassen  
REGISTRATION NUMBER: 31,845  
NAME: Donna E. Scherer  
REFERENCE/DOCKET NUMBER: CGNE 76-2 WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (916) 753-6313  
TELEFAX: (916) 753-1510  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2060 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
US-07-978-687-31

Query Match 1.0%; Score 19; DB 1; Length 2060;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1971 TTTTAAATAATAAAAAA 1989  
Db 113 TTTTAAATAATAAAAAA 131

RESULT 8  
PCT-US91-01750-2  
Sequence 2, Application PC/TUS9101750  
GENERAL INFORMATION:  
APPLICANT: KNAUF, VIC C.  
APPLICANT: KRIDL, JEAN C.  
APPLICANT: SCHERER, DONNA E.  
TITLE OF INVENTION: Novel Sequences Preferentially  
TITLE OF INVENTION: Expressed in Early Seed  
TITLE OF INVENTION: Development and Methods  
TITLE OF INVENTION: Related Thereto  
NUMBER OF SEQUENCES: 13  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calgene, Inc.  
STREET: 1920 Fifth Street  
CITY: Davis  
STATE: CA  
COUNTRY: USA  
ZIP: 95616  
COMPUTER-READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB storage  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/01750  
FILING DATE: 19910314  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/494,722  
FILING DATE: 16-MAR-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Elizabeth Lassen  
REGISTRATION NUMBER: 31,845  
NAME: Donna E. Scherer  
REFERENCE/DOCKET NUMBER: CGNE 68WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 916-753-6313  
TELEFAX: 916-753-1510  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2060 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
PCT-US91-01750-2

Query Match 1.0%; Score 19; DB 5; Length 2060;  
Best Local Similarity 100.0%; Pred. No. 39;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1971 TTTTAAATAATAAAAAA 1989  
Db 113 TTTTAAATAATAAAAAA 131

RESULT 9  
PCT-US91-05801-31  
Sequence 31, Application PC/TUS9105801  
GENERAL INFORMATION:  
APPLICANT: VIC. C. Knauf  
APPLICANT: Gregory A. Thompson  
TITLE OF INVENTION: Plant Patty Acid Synthases  
NUMBER OF SEQUENCES:  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Calgene, Inc.  
STREET: 1920 Fifth Street  
CITY: Davis  
STATE: CA  
COUNTRY: USA  
ZIP: 95616  
COMPUTER-READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB  
COMPUTER: Apple Macintosh  
OPERATING SYSTEM: Macintosh 6.0.7  
SOFTWARE: Microsoft Word 4.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/05801  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/568,493  
FILING DATE: 15-AUGUST-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/721,761  
FILING DATE: 26-JUNE-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Elizabeth Lassen  
REGISTRATION NUMBER: 31,845  
NAME: Donna E. Scherer  
REFERENCE/DOCKET NUMBER: CGNE 76-2 WO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (916) 753-6313  
TELEFAX: (916) 753-1510  
INFORMATION FOR SEQ ID NO: 31:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2060 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double

```
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
PCT-US91-05801-31

Query Match      1.0%; Score 19; DB 5; Length 2060;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1971 TTTTAAATAATAAAAAA 1989
Db 113 TTTTAAATAATAAAAAA 131

RESULT 10
PCT-US91-01750-3
; Sequence 3, Application PC/TUS9101750
; GENERAL INFORMATION:
; APPLICANT: KNAUF, VIC C.
; APPLICANT: KRIDL, JEAN C.
; TITLE OF INVENTION: Novel Sequences Preferentially
; TITLE OF INVENTION: Expressed In Early Seed
; TITLE OF INVENTION: Development and Methods
; TITLE OF INVENTION: Related Thereto
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB storage
; OPERATING SYSTEM: Macintosh 6.0
; SOFTWARE: Microsoft Word 4.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/01750
; FILING DATE: 19910314
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/494,722
; FILING DATE: 16-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; REFERENCE/DOCKET NUMBER: CGNE 68WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 916-753-6313
; TELEFAX: 916-753-1510
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2081 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
PCT-US91-01750-3

Query Match      1.0%; Score 19; DB 5; Length 2081;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1971 TTTTAAATAATAAAAAA 1989
Db 113 TTTTAAATAATAAAAAA 131

RESULT 11
US-08-471-791-27

; Sequence 27, Application US/08471791
; Patent No. 5723595
; GENERAL INFORMATION:
; APPLICANT: Thompson, Gregory A
; APPLICANT: Knauf, Vic C
; TITLE OF INVENTION: Plant Desaturases-Compositions
; TITLE OF INVENTION: and Uses
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: California
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
; OPERATING SYSTEM: Macintosh 7.1
; SOFTWARE: Microsoft Word 5.1 (a)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,791
; FILING DATE: 6-JUNE-95
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/762,762
; FILING DATE: 16-SEPT-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/01746
; FILING DATE: 14-MAR-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/615,784
; FILING DATE: 14-NOV-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/567,373
; FILING DATE: 13-AUG-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/494,106
; FILING DATE: 16-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Lassen, Elizabeth
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 69-5
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; TELEX: 350370 CGNE
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3440 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
US-08-471-791-27

Query Match      1.0%; Score 19; DB 1; Length 3440;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1971 TTTTAAATAATAAAAAA 1989
Db 344 TTTTAAATAATAAAAAA 362

RESULT 12
PCT-US91-01746-27
```



; Sequence 27, Application PC/TUS9101746  
; GENERAL INFORMATION:  
; APPLICANT: Thompson, Gregory A  
; APPLICANT: Knauf, Vic C  
; TITLE OF INVENTION: Plant Desaturases-Compositions and Uses  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Calgene, Inc.  
; STREET: 1920 Fifth Street  
; CITY: Davis  
; STATE: California  
; COUNTRY: USA  
; ZIP: 95616  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB storage  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: Macintosh 6.0.7  
; SOFTWARE: Microsoft Word 4.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US91/01746  
; FILING DATE: 19910314  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/615,784  
; FILING DATE: 14-NOV-1990  
; APPLICATION NUMBER: 07/567,373  
; FILING DATE: 13-AUG-1990  
; APPLICATION NUMBER: 07/494,106  
; FILING DATE: 16-MAR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Lassen, Elizabeth  
; REGISTRATION NUMBER: 31,845  
; NAME: Donna E. Scherer  
; REGISTRATION NUMBER: 34,719  
; REFERENCE/DOCKET NUMBER: CGNE 69-3 WO  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (916) 753-6313  
; TELEFAX: (916) 753-1510  
; TELEX: 350370 CGNE  
; INFORMATION FOR SEQ ID NO: 27:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3440 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
PCT-US91-01746-27  
  
Query Match 1.0%; Score 19; DB 5; Length 3440;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1971 TTTTAAATAATAAAAAA 1989  
Db 344 TTTTAAATAATAAAAAA 362  
  
RESULT 13  
US-09-220-132-51  
; Sequence 51, Application US/09220132  
; Patent No. 6506607  
; GENERAL INFORMATION:  
; APPLICANT: Sylvian, Andrew W.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT  
; FILE REFERENCE: 07334-074001  
; CURRENT APPLICATION NUMBER: US/09/220,132  
; CURRENT FILING DATE: 1998-12-23  
; PRIOR APPLICATION NUMBER: US 60/079,303  
; PRIOR FILING DATE: 1998-03-25  
; PRIOR APPLICATION NUMBER: US 60/068,821  
; PRIOR FILING DATE: 1997-12-24  
; NUMBER OF SEQ ID NOS: 191

; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 51  
; LENGTH: 3826  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-220-132-51  
  
Query Match 1.0%; Score 19; DB 4; Length 3826;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1974 TTAATAATAAAAAA 1992  
Db 3793 TTAATAATAAAAAA 3811  
  
RESULT 14  
US-08-737-078A-1/c  
; Sequence 1, Application US/08737078A  
; Patent No. 6027934  
; GENERAL INFORMATION:  
; APPLICANT: Powell, Curtis  
; TITLE OF INVENTION: VACCINE FOR, DIAGNOSTIC ASSAY FOR AND  
; METHOD OF TREATING PARASITIC HEMOFLAGELLATE PROTOZOA  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe  
; STREET: 660 White Plains Road  
; CITY: Tarrytown  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10591-5144  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage  
; COMPUTER: Apple Macintosh  
; OPERATING SYSTEM: System 7.5  
; SOFTWARE: Wordperfect 3.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/737,078A  
; FILING DATE: 16-JAN-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/04931  
; FILING DATE: 11-JUN-1992  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/714,464  
; FILING DATE: 13-JUN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/365,413  
; FILING DATE: 13-JUN-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kurt G. Briscoe  
; REGISTRATION NUMBER: 33,141  
; REFERENCE/DOCKET NUMBER: POWELL 203-PCT-KGB  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (914) 332-1700  
; TELEFAX: (914) 332-1844  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 141 nucleotides  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
US-08-737-078A-1  
  
Query Match 0.9%; Score 18; DB 3; Length 141;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1975 TAATAATAAAAAA 1992  
Db 67 TAATAATAAAAAA 50

## RESULT 15

PCT-US94-04706-1/c  
 ; Sequence 1, Application PC/TUS9404706  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Powell, Curtis  
 ; TITLE OF INVENTION: VACCINE FOR DIAGNOSTIC ASSAY FOR AND  
 ; TITLE OF INVENTION: METHOD OF TREATING PARASITIC  
 ; TITLE OF INVENTION: HEMOFLAGELLATE PROTOZOA  
 ; NUMBER OF SEQUENCES: 1  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Spring Horn Kramer & Woods  
 ; STREET: 660 White Plains Road  
 ; CITY: Tarrytown  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 10591-5144  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.50 inch, 800 kb storage  
 ; COMPUTER: Apple Macintosh  
 ; OPERATING SYSTEM: System 7.0  
 ; SOFTWARE: WordPerfect  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US94/04706  
 ; FILING DATE: Herewith  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US92/04931  
 ; FILING DATE: 11-JUN-92  
 ; PRIOR APPLICATION DATA: US 07/714,464  
 ; APPLICATION NUMBER: US 07/714,464  
 ; FILING DATE: 13-JUN-91  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 07/365,413  
 ; FILING DATE: 13-JUN-89  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kurt G. Briscoe  
 ; REGISTRATION NUMBER: 33,141  
 ; REFERENCE/DOCKET NUMBER: POWELL 203-PCT-KGB  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (914) 332-1700  
 ; TELEFAX: (914) 332-1844  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 141 nucleotides  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: double  
 ; TOPOLOGY: linear  
 ; PCT-US94-04706-1

Query Match 0.9%; Score 18; DB 5; Length 141;  
 Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1975 TAATAATAAAAAAAAAA 1992  
 Db 67 TAATAATAAAAAAAAAA 50

Search completed: January 31, 2004, 19:34:48  
 Job time : 128 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 13:58:54 ; Search time 7413 seconds  
(without alignments)  
10993.117 Million cell updates/sec

Title: US-09-744-313A-3

Perfect score: 1992

Sequence: 1-5tatgaaactccaaaggt.....tttaataataaaaaaaaaa 1992

Scoring table: OLIGO NUC

Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 20454813386 residues

Word size: 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.cm.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.in.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pln.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1992	100.0	1992	6	AX054819	Sequence
2	1993	80.0	1593	9	AF121863	AX054819 Homo sapi
3	1538	77.2	3038	9	BC005110	BC005110 Homo sapi
4	1369	68.7	2661	9	AY044865	AY044865 Homo sapi
5	1354	68.0	2925	9	BC046520	BC046520 Homo sapi
6	1315	66.0	3615	9	AK095380	AK095380 Homo sapi
7	1296	65.1	3145	6	AX512835	AX512835 Sequence
8	1296	65.1	3145	9	AK000362	AK000362 Homo sapi
9	1171	58.8	3576	9	AK026479	AK026479 Homo sapi
10	816	41.0	968	9	HS420561	AJ420561 Homo sapi
11	512	25.7	129010	9	AL589666	AL589666 Human DNA
12	425	21.3	425	6	BD113855	BD113855 EST and e
13	385	19.3	451	6	AX331384	AX331384 Sequence
14	357	17.9	358	6	BD026500	BD026500 Sequence
15	143	7.2	454	11	G30543	G30543 human STS S
16	81	4.1	1782	10	BC043328	BC043328 Mus muscu
17	78	3.9	580	6	AX387326	AX387326 Sequence
18	76	3.8	311	6	AX185131	AX185131 Sequence
19	76	3.8	313	6	AX186466	AX186466 Sequence
20	76	3.8	27898	2	AC116713	AC116713 Mus muscu
21	74	3.7	299	6	AX188027	AX188027 Sequence
22	74	3.7	224551	2	AC111832	AC111832 Rattus no
23	74	3.7	278375	2	AC130093	AC130093 Rattus no
24	35	1.8	160411	2	AC135935	AC135935 Rattus no
25	25	1.3	181833	2	AC141405	AC141405 Homo sapi
26	25	1.3	194034	9	AP005901	AP005901 Homo sapi
27	25	1.3	211492	2	AC140877	AC140877 Homo sapi
28	25	1.3	216811	2	AC140808	AC140808 Homo sapi
29	24	1.2	92628	9	AL355472	AL355472 Human DNA
30	24	1.2	145239	9	AC138701	AC138701 Homo sapi
31	24	1.2	149779	2	AL929118	AL929118 Homo sapi
32	24	1.2	150936	2	BX255899	BX255899 Danio rer
33	24	1.2	152668	5	AL935302	AL935302 Zebrafish
34	24	1.2	162533	2	EX537304	EX537304 Danio rer
35	24	1.2	168498	2	AL583847	AL583847 Homo sapi
36	24	1.2	169546	2	AC004157	AC004157 Plasmodi
37	24	1.2	174018	2	EX004989	EX004989 Danio rer
38	24	1.2	177789	2	EX470207	EX470207 Danio rer
39	24	1.2	179140	2	AC140860	AC140860 Homo sapi
40	24	1.2	179143	2	EX511195	EX511195 Danio rer
41	24	1.2	179577	5	AL928843	AL928843 Zebrafish
42	24	1.2	180715	2	EX005052	EX005052 Danio rer
43	24	1.2	217607	10	AL691509	AL691509 Mouse DNA
44	24	1.2	217698	2	EX248129	EX248129 Danio rer
45	24	1.2	224934	2	EX511028	EX511028 Danio rer

# ALIGNMENTS

RESULT 1	AX054819	AX054819	1992 bp	DNA	linear	PAT 13-JAN-2001
LOCUS	Sequence 3 from Patent WO0073334.					
DEFINITION	AX054819					
ACCESSION	AX054819.1	GI:12228268				
VERSION						
KEYWORDS	Homo sapiens (human)					
SOURCE	Homo sapiens					
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;					
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	Yue, H., Tang, Y.T. and Azimzai, Y.					
TITLE	Human sorting nexins					
JOURNAL	Patent: WO 0073334-A 3 07-DEC-2000;					

Incyte Genomics, Inc. (US)  
Location/Qualifiers  
1. 1992  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 671 a -308 c 372 g 641 t  
ORIGIN

Query Match 100.0%; Score 1992; DB 6; Length 1992;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATGAAACCTCAAAGTTATGTTAGCTATACCTTTAGTTTATCATTTTCAAACCTGTT 60  
DB 1 GTATGAAACCTCAAAGTTATGTTAGCTATACCTTTAGTTTATCATTTTCAAACCTGTT 60

QY 61 TTTCTTTTAAATGCTTTTCAATAAAGCTAGACTGTTGTTGATAAATTTGCTATGTA 120  
DB 61 TTTCTTTTAAATGCTTTTCAATAAAGCTAGACTGTTGTTGATAAATTTGCTATGTA 120

QY 121 CTTGATACATTTTGTGTTTAAATATTCAGGAACACACAGAAAGGGGAGAAATCATTTTGGAA 180  
DB 121 CTTGATACATTTTGTGTTTAAATATTCAGGAACACACAGAAAGGGGAGAAATCATTTTGGAA 180

QY 181 CAGCAGAAATAGTAGCAAAATTAAGAGATTAATCAAAAGTACCAATGAGAGGAGCTAT 240  
DB 181 CAGCAGAAATAGTAGCAAAATTAAGAGATTAATCAAAAGTACCAATGAGAGGAGCTAT 240

QY 241 GTTGCTTAATTTAGTGTAGCTGAAGTGAAGATGATTTTATTTGAAGAGGTTATGTTCT 300  
DB 241 GTTGCTTAATTTAGTGTAGCTGAAGTGAAGATGATTTTATTTGAAGAGGTTATGTTCT 300

QY 301 AATGGAAGATGATTTCCAGTGGAGGCTGTGAGCACACCTTAATCTCCCGGAAACCTTGC 360  
DB 301 AATGGAAGATGATTTCCAGTGGAGGCTGTGAGCACACCTTAATCTCCCGGAAACCTTGC 360

QY 361 TGCATGAAATATAGCAATTCATATGTAGACTTTTGTGAGGATCCCTCTCTGAAAGGAA 420  
DB 361 TGCATGAAATATAGCAATTCATATGTAGACTTTTGTGAGGATCCCTCTCTGAAAGGAA 420

QY 421 GGAGAAAGAGAGAGATTTCTGTTTGTATGATGTTGAAAGAAATGATAGAGAC 480  
DB 421 GGAGAAAGAGAGAGATTTCTGTTTGTATGTTGTTGTTGAAAGAAATGATAGAGAC 480

QY 481 AGTTGGACAGAGCTCAACATTTGCTGTCTATAGAGATATCTTGAATCTCTATGACT 540  
DB 481 AGTTGGACAGAGCTCAACATTTGCTGTCTATAGAGATATCTTGAATCTCTATGACT 540

QY 541 TGAATCAAAACTAACAGAAATTCATGTTGCAATTTCTGATGCCAGCTTCCTTCTAAGAG 600  
DB 541 TGAATCAAAACTAACAGAAATTCATGTTGCAATTTCTGATGCCAGCTTCCTTCTAAGAG 600

QY 601 GATCATTTGGCCCAAAATTAATGAAATTTAAAGTCAAGAGGAGAGTTCCAGAAATA 660  
DB 601 GATCATTTGGCCCAAAATTAATGAAATTTAAAGTCAAGAGGAGAGTTCCAGAAATA 660

QY 661 TCTACAGAACTCTGAGCATCCAGAACTGAGTAAGTCAACTCTGCGACATTTCT 720  
DB 661 TCTACAGAACTCTGAGCATCCAGAACTGAGTAAGTCAACTCTGCGACATTTCT 720

QY 721 TTCCCTTAATGGTGGGAAACACAAATTTCTTGATAGATACTACCAAGATTAATCTTGG 780  
DB 721 TTCCCTTAATGGTGGGAAACACAAATTTCTTGATAGATACTACCAAGATTAATCTTGG 780

QY 781 GAAATATATAAATCTGTTCTGAAATTAATGAAAGAGAAAGGTCAGCATTTGGAAC 840  
DB 781 GAAATATATAAATCTGTTCTGAAATTAATGAAAGAGAAAGGTCAGCATTTGGAAC 840

QY 841 TTTTATCATGATTTCAATTTCTGAGTCTCCAAAGCCTTAACCAAGTAGAGCCAGA 900  
DB 841 TTTTATCATGATTTCAATTTCTGAGTCTCCAAAGCCTTAACCAAGTAGAGCCAGA 900

QY 901 ACTGACCATTTCTCAGCCCTACTTTAGAAACCAACAAAGAGCTTTTCAATGATCTGTTAA 960  
DB 901 ACTGACCATTTCTCAGCCCTACTTTAGAAACCAACAAAGAGCTTTTCAATGATCTGTTAA 960

QY 961 AAATAATGCAAAACCGTCTGCAAAATACAGAGAGAAAGCAAAATCAAGATTTATTTATGGA 1020  
DB 961 AAATAATGCAAAACCGTCTGCAAAATACAGAGAGAAAGCAAAATCAAGATTTATTTATGGA 1020

QY 1021 GGTGATGACTGTAGAGAGCTCTATGATTTACCTGTATGTTAGGACGGTAGTTTCCA 1080  
DB 1021 GGTGATGACTGTAGAGAGCTCTATGATTTACCTGTATGTTAGGACGGTAGTTTCCA 1080

QY 1081 GGTTCCTGACTGGCTTCATCATCTCTTAATGGCAACTCGAATCTCTTTAAAAACACCTT 1140  
DB 1081 GGTTCCTGACTGGCTTCATCATCTCTTAATGGCAACTCGAATCTCTTTAAAAACACCTT 1140

QY 1141 GGAATGTATACCTGATTTACTATCTTCAGTGTAACTAGAACAGCTATTTTCAGAGACCG 1200  
DB 1141 GGAATGTATACCTGATTTACTATCTTCAGTGTAACTAGAACAGCTATTTTCAGAGACCG 1200

QY 1201 TTTGGTCTCACTCATTAACACTTCTCAGAGATGCTATATTTCTGTGAAAAACACTGAACTCG 1260  
DB 1201 TTTGGTCTCACTCATTAACACTTCTCAGAGATGCTATATTTCTGTGAAAAACACTGAACTCG 1260

QY 1261 CTCTCTCCAGATAAGCAAAAGAGCAAAAGAGCAAAAGAGCACTTTTGAAGAAATGATTAATCAT 1320  
DB 1261 CTCTCTCCAGATAAGCAAAAGAGCAAAAGAGCACTTTTGAAGAAATGATTAATCAT 1320

QY 1321 TCCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCACTGAGCTTCT 1380  
DB 1321 TCCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCACTGAGCTTCT 1380

QY 1381 GTTTGATGGCTTACAGCAACAGTACTCAACAGAGCTGACTTATGTTTATTTGGACAT 1440  
DB 1381 GTTTGATGGCTTACAGCAACAGTACTCAACAGAGCTGACTTATGTTTATTTGGACAT 1440

QY 1441 TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAGAGAGTACCTCTGTGAC 1500  
DB 1441 TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAGAGAGTACCTCTGTGAC 1500

QY 1501 ATCTTGATGTAAACACTTGGATTTGGTATAGATAACCAATTTGAAATTTCTGCTGTGCG 1560  
DB 1501 ATCTTGATGTAAACACTTGGATTTGGTATAGATAACCAATTTGAAATTTCTGCTGTGCG 1560

QY 1561 AGGCTGGTAGAAATTTACTTTTTGGGTATATTTCTTATATATATATATGTCATGCTGTC 1620  
DB 1561 AGGCTGGTAGAAATTTACTTTTTGGGTATATTTCTTATATATATATGTCATGCTGTC 1620

QY 1621 TGAATTTTGTAGTATTTTGTGTTTAAAGAGCTAAACAACTTTAATGATTAAGT 1680  
DB 1621 TGAATTTTGTAGTATTTTGTGTTTAAAGAGCTAAACAACTTTAATGATTAAGT 1680

QY 1681 GATTTGATCTCATAGTCTTTTCAATTTGCTAGTGTATCCAAATTTTATAGAACATTAAGT 1740  
DB 1681 GATTTGATCTCATAGTCTTTTCAATTTGCTAGTGTATCCAAATTTTATAGAACATTAAGT 1740

QY 1741 CACTTGTATTGCTCAATTTTAAAGAGAGAAATTCATTAATGATGTTATGCAAAACAGATAA 1800  
DB 1741 CACTTGTATTGCTCAATTTTAAAGAGAGAAATTCATTAATGATGTTATGCAAAACAGATAA 1800

QY 1801 GACTGATAAACTTCGTATTTGATAGCTTTGAAATAAATATGCTGATGATGAGAGAAACAG 1860  
DB 1801 GACTGATAAACTTCGTATTTGATAGCTTTGAAATAAATATGCTGATGATGAGAGAAACAG 1860

QY 1861 GAAATAGATCTGATTTTCTTAGAGTTAATATATTTTATAGTATGTTTCTTTTTTTT 1920  
DB 1861 GAAATAGATCTGATTTTCTTAGAGTTAATATATTTTATAGTATGTTTCTTTTTTTT 1920

QY 1921 ATTTTGTACATAGTTTAACTGTGATCTATAAATAAGCATCCCTATATGATGATTTTATAA 1980  
DB 1921 ATTTTGTACATAGTTTAACTGTGATCTATAAATAAGCATCCCTATATGATGATTTTATAA 1980

QY 1981 TAAAAAATAAATAA 1992



```
Db 1261 TAAAGACTAACAACTTAATGATTAAAGTGATTGAGTCTCATAGTCTTTTCATTGCT 1320
QY 1709 AGCTGATGATCCAAATTTATTAGACATAGTCACTGTTATTGCTATTTTAAAGAGA 1768
Db 1321 AGCTGATGATCCAAATTTATTAGACATAGTCACTGTTATTGCTATTTTAAAGAGA 1380
QY 1769 AAATTCATATGATGTTATGCGCAAAACAGATAAGACTGATAAATCTTCGTATTGATAGCTT 1828
Db 1381 AAATTCATATGATGTTATGCGCAAAACAGATAAGACTGATAAATCTTCGTATTGATAGCTT 1440
QY 1829 TGAATAATATATGCTAGTATGAGAGAAACAGATAAGACTGATAAATCTTCGTATTGATAGCTT 1888
Db 1441 TGAATAATATATGCTAGTATGAGAGAAACAGATAAGACTGATAAATCTTCGTATTGATAGCTT 1500
QY 1889 TATATTTTATGATGATGTTTCTTCTTTTATTTTATTTTATGATGATGATGATGATGATGAT 1948
Db 1501 TATATTTTATGATGATGTTTCTTCTTTTATTTTATTTTATGATGATGATGATGATGATGAT 1560
QY 1949 TAAATAAGCATCTATATGAGTATTTTAAATAAT 1981
Db 1561 TAAATAAGCATCTATATGAGTATTTTAAATAAT 1593

RESULT 3
BC005110 3038 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, clone MGC:13217 IMAGE:3959086, mRNA, complete cds.
DEFINITION BC005110
ACCESSION BC005110
VERSION BC005110.1 GI:13477272
KEYWORDS MGC.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 3038)
Direct Submission
Strausberg, R.
Submitted (26-MAR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NTH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@cgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
Schein, Duane Smaluis, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Nataaja van den Bosch, Jill Vardy,
George Yang, Scott Zynderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 18 Row: m Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.
FEATURES
Location/Qualifiers
1..3038
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:13217 IMAGE:3959086"
/tissue_type="Placenta, Chorioncarcinoma"
/clone_lib="NIH_MGC_21"
```

CDS

```
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
189..2849
/codon_start=1
/product="Unknown (protein for MGC:13217)"
/protein_id="AAH05110.1"
/db_xref="GI:13477273"
/translation="MQQKQRLRLDVGREICQVPLFCFLLCISAASLLNRYIHI
LMTFWGFAVVTFFYCSLQDLSLLPNIFTIKYKPKQLGLQELFPOGSCAVGKVKC
KEHPSSLLENVOPWLDLKITSSKVASLSEVDIPSIITKKLLPKKAMKHIEVIVKROK
VMTPELQAAALBEYGPPELHVALLSRDELHVLRLKLTLLPVLKPKATDCRSLLTLL
IRISLSVFLPSLDFADFTVNHLLIIFIDSPPEKATEPASELVPFLQKFAERNRN
KPSVLKLEQKIREQDLDFPMNFKQGAHVLFQCLTVEFDDRLIRLPELSNDE
MLSLHEELQKTYKTYCLIDBSIDKIRFDPITVEIQRARGPYIDVVKLQTMRLCFAY
EHLVSLLENVTFPMFCHSDYFQRLARGAESPTNSKLNKNTQKRGESGIGRIGKIS
KGVKSTMEGAMLPNVGAEGRDDFEISIVMEDDSPEVAVENTPRNLAAMKIS
IPVDFPEPSSERKEKRIIPVFCIDVERNDRRAVGHPEHMSVVRVLYLEFVLESK
LTFHGAFPAQPLPSKRIIGPKNYEFKSKREFQYELQKLOHPLELSQLLAFLS
PNGETQFLDKILPDVNLGKIISVFKLKRKGQHLPEFIMNFINSCESPKPKSRP
ELTILGPTSENKKLFNDLFKNNAEAENTKONQNYFMEVNTVEGVDPYKLVGRV
VPOVPDWLHLLMGTIRILPKNTLEMTDYLQCKLQELFQELHLSLILLLARDAIPCE
NTEPRSLQDKQKAKQTFEEMNYIPDLVKICIGETKYESIRLLFDGLQQPVNLQKL
TYVLLDIVIQELFPFLNKVQKEVTSVTSW"
```

BASE COUNT	949 a	564 c	642 g	883 t
Query Match	77.2%	Score 1538;	DB 9;	Length 3038;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1538;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	145	CAGGAACACACAGAAAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGGTAGCAAAATTA	204
Db <th>1481</th> <th>CAGGAACACACAGAAAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGGTAGCAAAATTA</th> <th>1540</th>	1481	CAGGAACACACAGAAAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGGTAGCAAAATTA	1540
QY <th>205</th> <th>AGGAGTATTCAAAGTACCAATGAGGAGGAGCTATGTTGCCTAATATTTGGTGTAGCTGA</th> <th>264</th>	205	AGGAGTATTCAAAGTACCAATGAGGAGGAGCTATGTTGCCTAATATTTGGTGTAGCTGA	264
Db <th>1541</th> <th>AGGAGTATTCAAAGTACCAATGAGGAGGAGCTATGTTGCCTAATATTTGGTGTAGCTGA</th> <th>1600</th>	1541	AGGAGTATTCAAAGTACCAATGAGGAGGAGCTATGTTGCCTAATATTTGGTGTAGCTGA	1600
QY <th>265</th> <th>AGGTGAGATGATTTTATTTGGAAGAGTATTTGTTGAATGGAAGATGATTTCCAGTGA</th> <th>324</th>	265	AGGTGAGATGATTTTATTTGGAAGAGTATTTGTTGAATGGAAGATGATTTCCAGTGA	324
Db <th>1601</th> <th>AGGTGAGATGATTTTATTTGGAAGAGTATTTGTTGAATGGAAGATGATTTCCAGTGA</th> <th>1660</th>	1601	AGGTGAGATGATTTTATTTGGAAGAGTATTTGTTGAATGGAAGATGATTTCCAGTGA	1660
QY <th>325</th> <th>GGCTGTGAGCACACTAATCTCTCCCGAAACCTTCTGTCATGGAATAATAGCATTCATA</th> <th>384</th>	325	GGCTGTGAGCACACTAATCTCTCCCGAAACCTTCTGTCATGGAATAATAGCATTCATA	384
Db <th>1661</th> <th>GGCTGTGAGCACACTAATCTCTCCCGAAACCTTCTGTCATGGAATAATAGCATTCATA</th> <th>1720</th>	1661	GGCTGTGAGCACACTAATCTCTCCCGAAACCTTCTGTCATGGAATAATAGCATTCATA	1720
QY <th>385</th> <th>TGTAGACTTTTTCAGGATCCCTCTCTGAAAGAGGAGAGAAAAAGAAATTCCTGT</th> <th>444</th>	385	TGTAGACTTTTTCAGGATCCCTCTCTGAAAGAGGAGAGAAAAAGAAATTCCTGT	444
Db <th>1721</th> <th>TGTAGACTTTTTCAGGATCCCTCTCTGAAAGAGGAGAGAAAAAGAAATTCCTGT</th> <th>1780</th>	1721	TGTAGACTTTTTCAGGATCCCTCTCTGAAAGAGGAGAGAAAAAGAAATTCCTGT	1780
QY <th>445</th> <th>GTTTGTATTCATGTTTGAAGAAATGATAGAGGAGGAGTTGGACAGGCTGACATTCG</th> <th>504</th>	445	GTTTGTATTCATGTTTGAAGAAATGATAGAGGAGGAGTTGGACAGGCTGACATTCG	504
Db <th>1781</th> <th>GTTTGTATTCATGTTTGAAGAAATGATAGAGGAGGAGTTGGACAGGCTGACATTCG</th> <th>1840</th>	1781	GTTTGTATTCATGTTTGAAGAAATGATAGAGGAGGAGTTGGACAGGCTGACATTCG	1840
QY <th>505</th> <th>GTCTGTCTATAGAGATATCTTGAATCTATGTTACTTGAATCAAACTCAACAGAAATTC</th> <th>564</th>	505	GTCTGTCTATAGAGATATCTTGAATCTATGTTACTTGAATCAAACTCAACAGAAATTC	564
Db <th>1841</th> <th>GTCTGTCTATAGAGATATCTTGAATCTATGTTACTTGAATCAAACTCAACAGAAATTC</th> <th>1900</th>	1841	GTCTGTCTATAGAGATATCTTGAATCTATGTTACTTGAATCAAACTCAACAGAAATTC	1900
QY <th>565</th> <th>TGGTGATTTTCTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATATGA</th> <th>624</th>	565	TGGTGATTTTCTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATATGA	624
Db <th>1901</th> <th>TGGTGATTTTCTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATATGA</th> <th>1960</th>	1901	TGGTGATTTTCTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATATGA	1960
QY <th>625</th> <th>ATTTCTTAAGTCAAGAGGAGGAGGAGTTCCAGAAATATCTACAGAACTTCTGCGAGCTCC</th> <th>684</th>	625	ATTTCTTAAGTCAAGAGGAGGAGGAGTTCCAGAAATATCTACAGAACTTCTGCGAGCTCC	684
Db <th>1961</th> <th>ATTTCTTAAGTCAAGAGGAGGAGGAGTTCCAGAAATATCTACAGAACTTCTGCGAGCTCC</th> <th>2020</th>	1961	ATTTCTTAAGTCAAGAGGAGGAGGAGTTCCAGAAATATCTACAGAACTTCTGCGAGCTCC	2020
QY <th>685</th> <th>AGACTGAGTAAATAGTCAACTTCTTGGCAGACTTCTTTTCCCTTAATGGTGGGGAACACA</th> <th>744</th>	685	AGACTGAGTAAATAGTCAACTTCTTGGCAGACTTCTTTTCCCTTAATGGTGGGGAACACA	744
Db <th>2021</th> <th>AGACTGAGTAAATAGTCAACTTCTTGGCAGACTTCTTTTCCCTTAATGGTGGGGAACACA</th> <th>2080</th>	2021	AGACTGAGTAAATAGTCAACTTCTTGGCAGACTTCTTTTCCCTTAATGGTGGGGAACACA	2080
QY <th>745</th> <th>ATTTCTTGTATAGATATCTACAGATGTAATCTTTGGGAAATATTAATAATCTGTTCTGG</th> <th>804</th>	745	ATTTCTTGTATAGATATCTACAGATGTAATCTTTGGGAAATATTAATAATCTGTTCTGG	804
Db <th>2081</th> <th>ATTTCTTGTATAGATATCTACAGATGTAATCTTTGGGAAATATTAATAATCTGTTCTGG</th> <th>2140</th>	2081	ATTTCTTGTATAGATATCTACAGATGTAATCTTTGGGAAATATTAATAATCTGTTCTGG	2140





[illegible]

RESULT	5
BC046520	
LOCUS	
DEFINITION	BC046520 Homo sapiens, Similar to sorting nexin 14, clone IMAGE:5267454, mRNA.
ACCESSION	BC046520
VERSION	BC046520.1 GI:28461364
	linear PRI 21-FEB-2003 2925 bp mRNA

KEYWORDS Homo sapiens (human)  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2925)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki  
Toshiyuki and Piero Carninci (RIKEN)  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [ang@bcm.tmc.edu](mailto:ang@bcm.tmc.edu)  
Gunsberg, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Lounsbury, H.,  
Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,  
A.N., Gibbs, R.A.  
Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAP Plate: 91 Row: a Column: 6.  
Location/Qualifiers  
1. .2925  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5267454"  
/tissue\_type="Testis"  
/clone\_lib="NIH MGC 97"  
/lab\_host="DH10B"  
/note="Vector: pBluescript"  
BASE COUNT 948 a 513 c 579 g 885 t  
ORIGIN  
Query Match 68.0%; Score 1354; DB 9; Length 2925;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1504; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 145 CAGGACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGTAGCAAAATTA 204  
DB 1380 CAGGACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGAAATAGTAGCAAAATTA 1439  
QY 205 AGGAGTATTCAGAAAGTACCAATGAGGGAGCTATGTCCTAATTCGTGTGAGCTGA 264  
DB 1440 AGGAGTATTCAGAAAGTACCAATGAGGGAGCTATGTCCTAATTCGTGTGAGCTGA 1499  
QY 265 AGGTGAAGATGATTTATTGAAGAAGGTATTGTTGTAATGGAAGATGATTCCTCAGTGA 324  
DB 1500 AGGTGAAGATGATTTATTGAAGAAGGTATTGTTGTAATGGAAGATGATTCCTCAGTGA 1559  
QY 325 GGCTGTGAGCACACCTAATACTCCCGAAACCTTGCTGCATGAGAAATAGCATTCATA 384  
DB 1560 GGCTGTGAGCACACCTAATACTCCCGAAACCTTGCTGCATGAGAAATAGCATTCATA 1619  
QY 385 TGTAGACTTTTGGAGATCCCTCCTCTGAAGGAGGAGAGAAAGAAAGAAATTCCTGT 444  
DB 1620 TGTAGACTTTTGGAGATCCCTCCTCTGAAGGAGGAGAGAGAAAGAAAGAAATTCCTGT 1679  
QY 445 GTTTTGTATTGATGTTGAAAGAAATGATAGAAGAGAGCTTTGGACACAGAGCTGAACATTG 504  
DB 1680 GTTTTGTATTGATGTTGAAAGAAATGATAGAAGAGAGCTTTGGACACAGAGCTGAACATTG 1739  
QY 505 GTCTGTCTATAGAAGATATCTTGAATTCATGACTTTGAATCAAACTAACAGAAATTTCA 564

1740 GTCTGTCTATAGAGATATCTTGAATTCATGATCTGATCAAACTAACAGAAATTC 1799  
565 TGGTGCATTTCTGATGCCAGCTTCTTCTTAAGAGATCATTTGGCCCAAAATTTATGA 624  
1800 TGGTGCATTTCTGATGCCAGCTTCTTCTTAAGAGATCATTTGGCCCAAAATTTATGA 1859  
625 ATTCCTTAAGTCAAGAGGAGAGTCCAGAGATATCTACAGAAATCTTCGAGCATCC 684  
1860 ATTCCTTAAGTCAAGAGGAGAGTCCAGAGATATCTACAGAAATCTTCGAGCATCC 1919  
685 AGAACTGAGTAAGTCAACTTCTCGGAGAGCTTTCTTCCCTTAATAGTGGGGAACACA 744  
1920 AGAACTGAGTAAGTCAACTTCTCGGAGAGCTTTCTTCCCTTAATAGTGGGGAACACA 1979  
745 ATTCCTTGAAGTCAACTTCTCGGAGAGCTTTCTTCCCTTAATAGTGGGGAACACA 804  
1980 ATTCCTTGAAGTCAACTTCTCGGAGAGCTTTCTTCCCTTAATAGTGGGGAACACA 2039  
805 AAAAATAAGTCAAGAGAGGTCAGCATTTGGAACTTTTATCATGAAATTCATTAATTC 864  
2040 AAAAATAAGTCAAGAGAGGTCAGCATTTGGAACTTTTATCATGAAATTCATTAATTC 2099  
865 TTGTGAGTCTCCAAAGCTTAAACCAAGTACAGCAACTGACCATTTCTCAGCCCTACTTC 924  
2100 TTGTGAGTCTCCAAAGCTTAAACCAAGTACAGCAACTGACCATTTCTCAGCCCTACTTC 2159  
925 AGAAAAACAAGAGAGCTTTCAATGATCTGTTTAAATAATTAATGCAACCGTCTGAAAA 984  
2160 AGAAAAACAAGAGAGCTTTCAATGATCTGTTTAAATAATTAATGCAACCGTCTGAAAA 2219  
985 TACAGAGAGAAACAAATGAAATTAATTTATGAGGTGATGATGAGAGAGTCTA 1044  
2220 TACAGAGAGAAACAAATGAAATTAATTTATGAGGTGATGATGAGAGAGTCTA 2279  
1045 TGATTACTGATGATGATGAGAGAGTATTTCCAGGTTCTGACTGCTTCTCATCTCT 1104  
2280 TGATTACTGATGATGATGAGAGAGTATTTCCAGGTTCTGACTGCTTCTCATCTCT 2339  
1105 CTTAATGGGAATCGAATCTCTTTAAACACACCGCTGGAATGTATGATGATCTCT 1164  
2340 CTTAATGGGAATCGAATCTCTTTAAACACACCGCTGGAATGTATGATGATCTCT 2399  
1165 TCAGTGTAAACTAGACAGCTATTTACAGGAGACCGTTTGGTCTCACTCATACACTTCT 1224  
2400 TCAGTGTAAACTAGACAGCTATTTACAGGAGACCGTTTGGTCTCACTCATACACTTCT 2459  
1225 CAGAGATGCTATATTTCTGTAACACTGAACTGCTCTCTCAGAGATGAGCAAAAGG 1284  
2460 CAGAGATGCTATATTTCTGTAACACTGAACTGCTCTCTCAGAGATGAGCAAAAGG 2519  
1285 AGCAAAACAGACTTTTGAAGAAATGAAATTAATTCAGATCTGTTAGTCAAGTGTAT 1344  
2520 AGCAAAACAGACTTTTGAAGAAATGAAATTAATTCAGATCTGTTAGTCAAGTGTAT 2579  
1345 TGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTGAGGCTTACAGCAACCACT 1404  
2580 TGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTGAGGCTTACAGCAACCACT 2639  
1405 ACTCAACAGCAGCTGATGATGTTTATTTGGAATGATGATGATGATGATGATGATGAT 1464  
2640 ACTCAACAGCAGCTGATGATGTTTATTTGGAATGATGATGATGATGATGATGATGAT 2699  
1465 GCTCAATAGGTACAAAGGAGTACCTCTGTGACATCTGGAATGATGATGATGATGATGAT 1524  
2700 GCTCAATAGGTACAAAGGAGTACCTCTGTGACATCTGGAATGATGATGATGATGATGAT 2759  
1525 TGGTATAGAAATTAACCCATTGAAATTTCTGTCGAGGAGGTTGATGAAATTTACTTTTTT 1584  
2760 TGGTATAGAAATTAACCCATTGAAATTTCTGTCGAGGAGGTTGATGAAATTTACTTTTTT 2819  
1585 GGGTATATTTCTATATATATATGATCGCTGCTGTAATTTTATGTTATTTTGTGTTT 1644

2820 GGGTATATTTCTATATATATATGATGATCGCTGCTGTAATTTTATGTTATTTTGTGTTT 2879  
1645 TTAATAA 1651  
2880 TTAATAA 2886  
RESULT 6  
AK095380  
LOCUS  
DEFINITION Homo sapiens CDNA FLJ38061 fis, clone CTONG2014966, highly similar to SORTING NEXIN 14.  
ACCESSION AK095380  
VERSION AK095380.1 GI:21754626  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Tanigami, A., Fujiwara, T., Shibahara, T., Goto, Y., Hirao, M., Shimizu, F., Wakebe, H., Ono, T., Hishigaki, H., Watanabe, T., Ozaki, K., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamaehita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Wagatsuma, N., Murakawa, K., Kanehori, K., Takahashi, Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEO human cDNA sequencing project  
Unpublished  
2 (bases 1 to 3616)  
Isogai, T. and Yamamoto, J.  
Direct Submission  
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 293-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
NEO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; CDNA full insert sequencing: Research Association for Biotechnology (RAB); CDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.  
FEATURES  
source  
1. 3616  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="CTONG2014966"  
/tissue type="tongue, tumor tissue"  
/clone\_lib="CTONG2"  
/note="Cloning vector: pME18SPL3"  
BASE COUNT 1078 a 629 c 710 g 1139 t  
ORIGIN  
Query Match 66.0%; Score 1315; DB 9; Length 3616;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1535; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
QY 145 CAGGAACACACAGAAAGGGGAGGAGTCAATTTGGATCAGCAGATAGGTAGCAAAATTA 204  
DB 2079 CAGGAACACACAGAAAGGGGAGGAGTCAATTTGGATCAGCAGATAGGTAGCAAAATTA 2138  
QY 205 AGGAGTATTTCAAAGTACCACAAATGGAGGAGCTATGTTGCCTAATTTATGTTAGCTGA 264  
DB 2139 AGGAGTATTTCAAAGTACCACAAATGGAGGAGCTATGTTGCCTAATTTATGTTAGCTGA 2198  
QY 265 AGGTCAACATGATTTTATTTGAAGAGGTATTTGTTGTAATGGAAGATGATTTCTCCAGTGA 324  
DB 2199 AGGTCAACATGATTTTATTTGAAGAGGTATTTGTTGTAATGGAAGATGATTTCTCCAGTGA 2258  
QY 325 GGCTGTGAGCACACCTAATACTCCCGAAACCTTGTGCTGCATGGAAATTAATGATTTCCATA 384

Db 2259 GGCTGTGAGCACACCTTAATACCTCCCGAAACCTTGTGTCATGGAATTTAGCATTTCCATA 2318  
Qy TGTAGACTTTTTTGGAGTCCCTCTCTGAAAGGAGGAG-AAAAGAAAGAAATTCCTG 443  
Db 2319 TGTAGACTTTTTTGGAGTCCCTCTCTGAAAGGAGGAG-AAAAGAAAGAAATTCCTG 2378  
Qy 444 TGTGTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCCTGACACATT 503  
Db 2379 TGTGTTTGTATGATGTTGAAAGAAATGATAGAGAGCAGTTGGACACGAGCCTGACACATT 2438  
Qy 504 GGTCTGTCTATAGAGATCTTTGAAATCTATGATCTTGAAATCAAACTAAACAGAAATTC 563  
Db 2439 GGTCTGTCTATAGAGATCTTTGAAATCTATGATCTTGAAATCAAACTAAACAGAAATTC 2498  
Qy 564 ATGGTGCATTTCTGTATGCCAGCTTCCCTCTTAAGAGGATCATTTGGCCCCCAAAATATG 623  
Db 2499 ATGGTGCATTTCTGTATGCCAGCTTCCCTCTTAAGAGGATCATTTGGCCCCCAAAATATG 2558  
Qy 624 AATCTTTAAAGTCAAGAGGAGAGATTCCAAGATATCTACAGAACTTCTGCAGCATC 683  
Db 2559 AATCTTTAAAGTCAAGAGGAGAGATTCNAGAAATATCTACAGAACTTCTGCAGCATC 2618  
Qy 684 CAGAACTGAGTAATAGTCAACTTCTGCAGACTTTCTTTCCCTAAATGTTGGGAAACAC 743  
Db 2619 CAGAACTGAGTAATAGTCAACTTCTGCAGACTTTCTTTCCCTAAATGTTGGGAAACAC 2678  
Qy 744 AATTTCTGTATAGATCTACAGATGTAATCTTGGGAAATATTAATCTGTTCTG 803  
Db 2679 AATTTCTGTATAGATCTACAGATGTAATCTTGGGAAATATTAATCTGTTCTG 2738  
Qy 804 GAAACTTAATGAAGAGAGAGTTCAGATTTGAACTTTTATCATGAATTTTCAATTAAT 863  
Db 2739 GAAACTTAATGAAGAGAGAGTTCAGATTTGAACTTTTATCATGAATTTTCAATTAAT 2798  
Qy 864 CTGTGTAGTCTCCAAAGCTTAACCAAGTAGACAGAACTGACCATCTCAGCCCTACTT 923  
Db 2799 CTGTGTAGTCTCCAAAGCTTAACCAAGTAGACAGAACTGACCATCTCAGCCCTACTT 2858  
Qy 924 CAGAAACAAACAGAGCTTTCAATGATCTGTTTAAATAATGCAAAACCGTGTGAAA 983  
Db 2859 CAGAAACAAACAGAGCTTTCAATGATCTGTTTAAATAATGCAAAACCGTGTGAAA 2918  
Qy 984 ATACAGAGAGAGAGAGAGTTCAGATTTTATGAGGAGTATGATCTGAGAGGAGTCT 1043  
Db 2919 ATACAGAGAGAGAGAGAGTTCAGATTTTATGAGGAGTATGATCTGAGAGGAGTCT 2978  
Qy 1044 ATGATTACCTGATCTATGTAGAGCGGTAGTTTTCAGAGTTCCTGACTGGCTTCATCATC 1103  
Db 2979 ATGATTACCTGATCTATGTAGAGCGGTAGTTTTCAGAGTTCCTGACTGGCTTCATCATC 3038  
Qy 1104 TCTTAATGGGAACTCGAATCCTTTTAAACACACCTTGGAAATGATATCTGATTAATC 1163  
Db 3039 TCTTAATGGGAACTCGAATCCTTTTAAACACACCTTGGAAATGATATCTGATTAATC 3098  
Qy 1164 TTGAGTGAAGTGAAGAGTATTTAGAGGAGCGGTGTTGGTCTCACTATTAACACTTC 1223  
Db 3099 TTGAGTGAAGTGAAGAGTATTTAGAGGAGCGGTGTTGGTCTCACTATTAACACTTC 3158  
Qy 1224 TCAGAGATGCTATATCTGTGAAACACACTGAACTCTCTCCAGAGTAAAGAAAAG 1283  
Db 3159 TCAGAGATGCTATATCTGTGAAACACACTGAACTCTCTCCAGAGTAAAGAAAAG 3218  
Qy 1284 GAGCAAAACAGACTTTTGAAGAAATGATGAATTTACATTCAGATCTGTTAGTCAAGTGA 1343  
Db 3219 GAGCAAAACAGACTTTTGAAGAAATGATGAATTTACATTCAGATCTGTTAGTCAAGTGA 3278  
Qy 1344 TTGFTGAAGAACCAAGATGAAAGCATTCAGACTTTCTGTTTGTATGAGTTCAGCAACAG 1403  
Db 3279 TTGFTGAAGAACCAAGATGAAAGCATTCAGACTTTCTGTTTGTATGAGTTCAGCAACAG 3338  
Qy 1404 TACTCAACAGAGCTGATCTATGTTTATTTGAGATGATGATACAGAACTGTTCCAG 1463

Db 3339 TACTCAACAGAGAGCTGACTTATGTTTATTGACATGATGATACAGGAACTGTTTCCAG 3398  
Qy 1464 AGCTCAATAAGGTACAAAAGAAAGTTTACCTCTGTGACATCTTGGATGTAAACAATTGGAT 1523  
Db 3399 AGCTCAATAAGGTACAAAAGAAAGTTTACCTCTGTGACATCTTGGATGTAAACAATTGGAT 3458  
Qy 1524 TTGGTATAGAAATTAACCCATTTGAATTTCTGCTGTGCGAGGTTGTAGAAATTTACTTTT 1583  
Db 3459 TTGGTATAGAAATTAACCCATTTGAATTTCTGCTGTGCGAGGTTGTAGAAATTTACTTTT 3518  
Qy 1584 TTGGTATATTTCTTATATATATATATGATGATCATGCTGTCTGAAATTTTAGTTTATTTTGT 1643  
Db 3519 TTGGTATATTTCTTATATATATATATGATGATCATGCTGTCTGAAATTTTAGTTTATTTTGT 3578  
Qy 1644 TTTAATAAAGACTAAACAACTTAATGATTTAAAGTG 1681  
Db 3579 TTTAATAAAGACTAAACAACTTAATGATTTAAAGTG 3616

RESULT 7  
AX512835  
LOCUS AX512835 3145 bp DNA linear PAT 03-OCT-2002  
DEFINITION Sequence 4 from Patent WO02062839.  
ACCESSION AX512835  
VERSION AX512835.1 GI:23504019  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1  
Daemen, M.J., Cleutjens, C.B. and Zaman, G.J.  
Markers of unstable atherosclerotic plaques  
Patent: WO 02062839-A 4 15-AUG-2002;  
Universiteit Maastricht (NL)  
Location/Qualifiers  
1. 3145  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 985 a 570 c 659 g 931 t  
ORIGIN

Query Match 65.1%; Score 1296; DB 6; Length 3145;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1446; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 147 GGAACACACAGAAAGGGGAGATCATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 206  
Db 1602 GGAACACACAGAAAGGGGAGATCATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 1661  
Qy 207 GAGTATTCAAAAGTACCAATGAGGAGCTATTTGCCCTAATTTATGTTAGCTGAG 266  
Db 1662 GAGTATTCAAAAGTACCAATGAGGAGCTATTTGCCCTAATTTATGTTAGCTGAG 1721  
Qy 267 GTGAGATGATTTTATGAGAGGATTTGTTGAAATGGAAGATGTTCTCCAGTGAGG 326  
Db 1722 GTGAGATGATTTTATGAGAGGATTTGTTGAAATGGAAGATGTTCTCCAGTGAGG 1781  
Qy 327 CTGTGAGCAGACCTAAATACTCCCGAAACCTTGTGCTGATGAAATTTAGCAATTCATG 386  
Db 1782 CTGTGAGCAGACCTAAATACTCCCGAAACCTTGTGCTGATGAAATTTAGCAATTCATG 1841  
Qy 387 TAGACTTTTTTGAAGATCCCTCTCTGAAAGAGGAGGAAAGAAAGAAATTTCTGTGT 446  
Db 1842 TAGACTTTTTTGAAGATCCCTCTCTGAAAGAGGAGGAAAGAAAGAAATTTCTGTGT 1901  
Qy 447 TTTGATTTGATTTGAAAGAAATGATAGAGAGAGTTGGACAGCCTGAAACATTTGT 506  
Db 1902 TTTGATTTGATTTGAAAGAAATGATAGAGAGAGTTGGACAGCCTGAAACATTTGT 1961  
Qy 507 CTGTCTATAGAGATATCTGAAATTTCTATGTTGATCAAACTAAACAAATTTTCTATG 566

Db 1962 CTGCTATAGAGATATCTTGAATTCATGTACTGAATCAAACTAACAGAAATTTTCATG 2021  
Qy 567 GTGCATTTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATTTATGAAT 626  
Db 2022 GTGCATTTCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCCCAAAATTTATGAAT 2081  
Qy 627 TCTTAAAGTCAAGAGAGGAGAGTTCAGAAATATCTACAGAAACTTCTGCAGCATCCAG 686  
Db 2082 TCTTAAAGTCAAGAGAGGAGAGTTCAGAAATATCTACAGAAACTTCTGCAGCATCCAG 2141  
Qy 687 AACTGAGTAATAGTCAACTTCTGCGAGACTTCTTCCCTTAAATGTGGGGAAACACAT 746  
Db 2142 AACTGAGTAATAGTCAACTTCTGCGAGACTTCTTCCCTTAAATGTGGGGAAACACAT 2201  
Qy 747 TCTTGTAAGTACTACAGAGTAAATCTTGGAAATTAATAATCTGTCTCTGGAA 806  
Db 2202 TCTTGTAAGTACTACAGAGTAAATCTTGGAAATTAATAATCTGTCTCTGGAA 2261  
Qy 807 AACTTAATGAAGAGAGAGTTCAGCAATTTGGAACCTTTTATCATGAATTTCAATATCTT 866  
Db 2262 AACTTAATGAAGAGAGAGTTCAGCAATTTGGAACCTTTTATCATGAATTTCAATATCTT 2321  
Qy 867 GTGAGTCTCAAAGCCTTAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTCAG 926  
Db 2322 GTGAGTCTCAAAGCCTTAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTCAG 2381  
Qy 927 AAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGAAATA 986  
Db 2382 AAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGAAATA 2441  
Qy 987 CAGAGAGAGAGCAAAATCAGAAATTTATGAGAGTGATGACTGAGAGAGTCTATG 1046  
Db 2442 CAGAGAGAGAGCAAAATCAGAAATTTATGAGAGTGATGACTGAGAGAGTCTATG 2501  
Qy 1047 ATTACCTGATGTATGAGAGCGGTAGTTTCCAGGTTCTCTGAGTGGCTTCATCTCT 1106  
Db 2502 ATTACCTGATGTATGAGAGCGGTAGTTTCCAGGTTCTCTGAGTGGCTTCATCTCT 2561  
Qy 1107 TAATGGGAATCGAATCTCTTTTAAACACCCCTGGAATGTATGATTAATCTATCTTC 1166  
Db 2562 TAATGGGAATCGAATCTCTTTTAAACACCCCTGGAATGTATGATTAATCTATCTTC 2621  
Qy 1167 AGTGTAACTAGAACAGCTATTTTCAGAGCACCGTTTGGTCTCACTCATAACTCTCA 1226  
Db 2622 AGTGTAACTAGAACAGCTATTTTCAGAGCACCGTTTGGTCTCACTCATAACTCTCA 2681  
Qy 1227 GAGATGCTATTTCTGTAAGAACACTGAACTCGCTCTCTCCAGATTAAGCAAAAGGAG 1286  
Db 2682 GAGATGCTATTTCTGTAAGAACACTGAACTCGCTCTCTCCAGATTAAGCAAAAGGAG 2741  
Qy 1287 CAACACAGACTTTTGAAGAAATGATGAATTCAGATCTGTTAGTCAAGTGATTTG 1346  
Db 2742 CAACACAGACTTTTGAAGAAATGATGAATTCAGATCTGTTAGTCAAGTGATTTG 2801  
Qy 1347 GTGAGAAACCAAGTATGAAGCATCGACTTCTGTTTGAATGCTTACAGCAACCGATAC 1406  
Db 2802 GTGAGAAACCAAGTATGAAGCATCGACTTCTGTTTGAATGCTTACAGCAACCGATAC 2861  
Qy 1407 TCACAGAGAGCTGACTTATGTTTATTTGACATTTGATACAGGAACCTGTTTCCAGAG 1466  
Db 2862 TCACAGAGAGCTGACTTATGTTTATTTGACATTTGATACAGGAACCTGTTTCCAGAG 2921  
Qy 1467 TCATATAGGTACAAAGAGAGTTCACCTCTGTGACATCTTGGATGAACACTTGGATTTG 1526  
Db 2922 TCATATAGGTACAAAGAGAGTTCACCTCTGTGACATCTTGGATGAACACTTGGATTTG 2981  
Qy 1527 GTATAGAAATACCCATTGAAATTTCTGCTGTGCGAGGGTGGTGAATAATTTACTTTTTCG 1586  
Db 2982 GTATAGAAATACCCATTGAAATTTCTGCTGTGCGAGGGTGGTGAATAATTTACTTTTTCG 3041  
Qy 1587 GTATATTTCT 1595  
Db 3042 GTATATTTCT 3050

## RESULT 8

AK000362

## LOCUS

## DEFINITION

AK000362

## ACCESSION

AK000362

## VERSION

AK000362.1

## KEYWORDS

oligo capping; fis (full insert sequence).

## SOURCE

Homo sapiens

## ORGANISM

Homo sapiens

## REFERENCE

1 (sites)

## AUTHORS

Kawakami,T., Noguchi,S., Itoh,T., Shigeta,K., Senba,T.,

Matsumura,K., Nakajima,Y., Mizuno,T., Morinaga,M., Ota,T.,

Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,

Nakamura,Y., Isogai,T. and Sugano,S.

## TITLE

NEBO human cDNA sequencing project

## JOURNAL

Unpublished

## AUTHORS

2 (bases 1 to 3145)

Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,

Shibahara,T., Tanaka,T. and Nakamura,Y.

## TITLE

Direct Submission

## JOURNAL

Submitted (15-FEB-2000)

Sumio Sugano, Institute of Medical Science,

University of Tokyo, Department of Virology, Shirokane-dai, 4-6-1,

Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp,

Tel:81-3-5449-5286, Fax:81-3-5449-5416)

## COMMENT

NEBO human cDNA sequencing project supported by Ministry of

International Trade and Industry of Japan; cDNA full insert

sequencing: Research Association for Biotechnology; cDNA library

construction, 5'- &amp; 3'-end one pass sequencing: Department of

Virology and Human Genome Center, Institute of Medical Science,

University of Tokyo (partly supported by Science and Technology

Agency).

## FEATURES

Location/Qualifiers

1..3145

/organism="Homo sapiens"

/mol\_type="mRNA"

/db\_xref="taxon:9606"

/clones="HEP15804"

/cell\_line="HepG2"

/cell\_type="hepatoma"

/clone\_lib="HEP"

/note="cloning vector pME18SFL3"

misc\_feature

1..3145

/note="highly similar to AF121863 Homo sapiens sorting

nexin 14"

BASE COUNT

985 a 570 c 659 g 931 t

ORIGIN

Query Match 65.1%; Score 1296; DB 9; Length 3145;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1446; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 147 GGAACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 206

Db 1602 GGAACACACAGAAAGGGGAGAAATCATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 1661

Qy 207 GAGTATTCAGAAAGTACCAATGGAGGAGCTATGTTGCCCTAATATTGTTAGTCTGAG 266

Db 1662 GAGTATTCAGAAAGTACCAATGGAGGAGCTATGTTGCCCTAATATTGTTAGTCTGAG 1721

Qy 267 GTGAGATGATTTTATTCAGAGAGGTATTTGTTAATGGAAGATGATTTCCAGTGGAGG 326

Db 1722 GTGAGATGATTTTATTCAGAGAGGTATTTGTTAATGGAAGATGATTTCCAGTGGAGG 1781

Qy 327 CTGTGAGCACACCTTAATCTCCCGAAACCTTGTGATCGTGAATAATAGCAATTCATATG 386

Db 1782 CTGTGAGCACACCTTAATCTCCCGAAACCTTGTGATCGTGAATAATAGCAATTCATATG 1841

Qy 387 TAGACTTTTTTGAGGATCCCTCTCTGAAAGGAGGAGAGAGAGAGAGAGAGAGAGAGAG 446





QY 265 AGGTGAAGTGAATTTATTTGAGAGAGGTAATGTTGTAATGGAAGATGATCTCCAGTGA 324  
Db 2136 AGGTGAAGTGAATTTATTTGAGAGAGGTAATGTTGTAATGGAAGATGATCTCCAGTGA 2195  
QY 325 GGCTGTGAGCACACCTTAATCTCCGAAACCTTGTGTCATCGAAATTTAGCATTTCCATA 384  
Db 2196 GGCTGTGAGCACACCTTAATCTCCGAAACCTTGTGTCATCGAAATTTAGCATTTCCATA 2255  
QY 385 TGTAAGACTTTTTTGAGGATCCCTCTCTGAAAGGAGGAGAGAAAAGAAAGAAATTCCTGT 444  
Db 2256 TGTAAGACTTTTTTGAGGATCCCTCTCTGAAAGGAGGAGAGAAAAGAAAGAAATTCCTGT 2315  
QY 445 GTTTGTGATGATGTTGAAAGAAATGATAGAGAGAGTTGGACACGAGCCTGAAACATTTG 504  
Db 2316 GTTTGTGATGATGTTGAAAGAAATGATAGAGAGAGTTGGACACGAGCCTGAAACATTTG 2375  
QY 505 GTCTGCTATAGAGATATCTTGAATCTTATGATCTTGAATCAAACTAACAGATTTCA 564  
Db 2376 GTCTGCTATAGAGATATCTTGAATCTTATGATCTTGAATCAAACTAACAGATTTCA 2435  
QY 565 TGTGTGATTTCCCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCAAAAATTTATGA 624  
Db 2436 TGTGTGATTTCCCTGATGCCAGCTTCTCTTAAGAGGATCATTTGGCCCAAAAATTTATGA 2495  
QY 625 ATTCCTTAAAGTCAAGAGGGAAGAGTTCCAAAGATATCTACAGAAACTTCTCAGCATCC 684  
Db 2496 ATTCCTTAAAGTCAAGAGGGAAGAGTTCCAAAGATATCTACAGAAACTTCTCAGCATCC 2555  
QY 685 AGAAGTGAATAGTCAACTTCTGGCAGACTTCTTCCCTTAAGGTTGGGGGAAACACA 744  
Db 2556 AGAAGTGAATAGTCAACTTCTGGCAGACTTCTTCCCTTAAGGTTGGGGGAAACACA 2615  
QY 745 ATTTCTTTGATAGATATCTACAGATGTAATCTTGGGAAATTTATAAATCTGTTCTCG 804  
Db 2616 ATTTCTTTGATAGATATCTACAGATGTAATCTTGGGAAATTTATAAATCTGTTCTCG 2675  
QY 805 AAAAATAAGAGAGAGAGGCTCAGCATTTTGAACCTTTTATCATGAAATTTCAATTAATTC 864  
Db 2676 AAAAATAAGAGAGAGAGGCTCAGCATTTTGAACCTTTTATCATGAAATTTCAATTAATTC 2735  
QY 865 TTGTGATCTCCAAAGCCTAAACCAAGTAGACAGAACTGACATTTCTCAGCCTTACTTC 924  
Db 2736 TTGTGATCTCCAAAGCCTAAACCAAGTAGACAGAACTGACATTTCTCAGCCTTACTTC 2795  
QY 925 AGAAAAACAAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAAA 984  
Db 2796 AGAAAAACAAGAGAGCTTTTCAATGATCTGTTTAAATAATGCAACCGTCTGAAA 2855  
QY 985 TACAGAGAGAGAGAGAGAGGATTTTATGAGGATGATGATGATGATGATGATGATGATGAT 1044  
Db 2856 TACAGAGAGAGAGAGAGAGGATTTTATGAGGATGATGATGATGATGATGATGATGATGAT 2915  
QY 1045 TGATTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1104  
Db 2916 TGATTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2975  
QY 1105 CTTAATGGGAACTGGAATCTCTTTTAAACACACCTGGAATGATGATGATGATGATGATGAT 1164  
Db 2976 CTTAATGGGAACTGGAATCTCTTTTAAACACACCTGGAATGATGATGATGATGATGATGAT 3035  
QY 1165 TCAGTGTAACTAGAACAGCTATTTTACGAGGACACCGTTTGGTCTCACTCATACATTTCT 1224  
Db 3036 TCAGTGTAACTAGAACAGCTATTTTACGAGGACACCGTTTGGTCTCACTCATACATTTCT 3095  
QY 1225 CAGAGATGCTATATTTCTGTGAAACACTGAACTCTGCTCTCTCAAGATGAGCAAAAGG 1284  
Db 3096 CAGAGATGCTATATTTCTGTGAAACACTGAACTCTGCTCTCTCAAGATGAGCAAAAGG 3155  
QY 1285 AGCAAAACAGACTTTTGAAGAAATGATGAAATTTACATTTCCAGATCTGTTAGTCAAGTGTAT 1344  
Db 3156 AGCAAAACAGACTTTTGAAGAAATGATGAAATTTACATTTCCAGATCTGTTAGTCAAGTGTAT 3215

QY 1345 TGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTGATGGCTTACAGCAACCACT 1404  
Db 3216 TGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTGATGGCTTACAGCAACCACT 3275  
QY 1405 ACTCAACAAGCAG 1417  
Db 3276 ACTCAACAAGCAG 3288  
RESULT 10  
LOCUS HSA420561 968 bp mRNA linear PRI 23-NOV-2001  
DEFINITION Homo sapiens mRNA full length insert cdna clone EUROIIMAGE 1521079.  
ACCESSION AJ420561  
VERSION AJ420561.1 GI:17066425  
KEYWORDS human gene transcripts  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1  
AUTHORS Auffray, C., Ansong, W., Ballabio, A., Estivill, X., Gibson, K.,  
Lehrach, H., Poustka, A. and Lundberg, J.  
TITLE The European IMAGE consortium for integrated Molecular analysis of  
human gene transcripts  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 968)  
AUTHORS Persson, A.  
TITLE Direct Submission  
JOURNAL Submitted (02-OCT-2001) Persson A., Center for Molecular  
Biotechnology, KTH, SCFAB, Institute of Biotechnology, Roslagsvagen  
30B, 106 91 Stockholm, SWEDEN  
COMMENT This clone is available royalty-free through IMAGE Consortium  
Distributors. IMPORTANT: This sequence represents the full insert  
of this IMAGE cdna clone. No attempt has been made to verify  
whether this corresponds to the full-length of the original mRNA  
from which it was derived.  
FEATURES  
source  
1. 968  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/chromosome="6"  
/clones="EUROIIMAGE 1521079"  
/clone\_lib="NCI CGAP Lu5"  
BASE COUNT 311 a 147 c 186 g 324 t  
ORIGIN  
Query Match 41.0%; Score 816; DB 9; Length 968;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 936; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 1021 GGTGATGATCTGAGAGGAGTCTATGATTACCTGATGATGATGATGATGATGATGATGAT 1080  
Db 20 GGTGATGATCTGAGAGGAGTCTATGATTACCTGATGATGATGATGATGATGATGATGAT 79  
QY 1081 GGTCTCTGACCTGGCTTCATCTCTTAATGGGAACCTGAACTCTCTTTAAAAACACCT 1140  
Db 80 GGTCTCTGACCTGGCTTCATCTCTTAATGGGAACCTGAACTCTCTTTAAAAACACCT 139  
QY 1141 GGAATGTATATCTGATTAATCTCTTCTGATGATGATGATGATGATGATGATGATGAT 1200  
Db 140 GGAATGTATATCTGATTAATCTCTTCTGATGATGATGATGATGATGATGATGATGAT 199  
QY 1201 TTTGTTCTCACTCAATAACCTTCTCAGAGATGCTATATTTCTGTGAAAACACTGAACTTCG 1260  
Db 200 TTTGTTCTCACTCAATAACCTTCTCAGAGATGCTATATTTCTGTGAAAACACTGAACTTCG 259  
QY 1261 CTCTCTCCAGATGAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGATGATGAT 1320  
Db 260 CTCTCTCCAGATGAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGATGATGAT 319  
QY 1321 TCCAGATCTGTTAGTCAAGTGTATTGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCT 1380

```

Db 320 TCCAGATCTCTAGTCAAGTGTTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCT 379
Qy 1381 GTTTGATGGTTACAGCAACAGTACTCAACAGCAGCTGACTATGTTTATTGACAT 1440
Db 380 GTTTGATGGTTACAGCAACAGTACTCAACAGCAGCTGACTATGTTTATTGACAT 439
Qy 1441 TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAAGGAAGTTACCTCTGTGAC 1500
Db 440 TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAAGGAAGTTACCTCTGTGAC 499
Qy 1501 ATCTTGATGTAACACTTGGATTGGTATAGTAATACCCATTGAAATTTCTGCTGTGG 1560
Db 500 ATCTTGATGTAACACTTGGATTGGTATAGTAATACCCATTGAAATTTCTGCTGTGG 559
Qy 1561 AGGTGTGAGAAATTTACTTTTGGTATATTTCTTATATATATATATATATATATATATAT 1620
Db 560 AGGTGTGAGAAATTTACTTTTGGTATATTTCTTATATATATATATATATATATATATAT 619
Qy 1621 TGAATTTTGTATATTTTTT-GTTTTTAAATAGAGCTAACACAACTTAATGATTAATAAG 1679
Db 620 TGAATTTTGTATATTTTTTGGTTTTTAAATAGAGCTAACACAACTTAATGATTAATAAG 679
Qy 1680 TGATTGAGTCTCATAGTCTTTTCATTTCTAGCTGTGATCCAAATTTTATTAGAACATAAG 1739
Db 680 TGATTGAGTCTCATAGTCTTTTCATTTCTAGCTGTGATCCAAATTTTATTAGAACATAAG 739
Qy 1740 TCATTGTTATGTCATTTTAAAGAGAAATTCATATGATGTTATGTCGAAACAGATA 1799
Db 740 TCATTGTTATGTCATTTTAAAGAGAAATTCATATGATGTTATGTCGAAACAGATA 799
Qy 1800 AGACTGATAAATTCGATTTGTATAGCTTTGAAAATAATATATCCCTAGTATGAGAAACA 1859
Db 800 AGACTGATAAATTCGATTTGTATAGCTTTGAAAATAATATATCCCTAGTATGAGAAACA 859
Qy 1860 GGAATAAGATCTGATTTTCTTAGAGTAAATATATTTTAGTAGATGTTGTTTCTCTTTT 1919
Db 860 GGAATAAGATCTGATTTTCTTAGAGTAAATATATTTTAGTAGATGTTGTTTCTCTTTT 919
Qy 1920 TATTTGTACATGATTAAGTGTGATCTATAATAATAA 1956
Db 920 TATTTGTACATGATTAAGTGTGATCTATAATAATAA 956

```

```

RESULT 11
AL589666/c 129010 bp DNA linear PRI 12-MAY-2001
LOCUS Human DNA sequence from clone RP11-321N4 on chromosome 6, complete
DEFINITION sequence.
ACCESSION AL589666
VERSION AL589666.5 GI:14041764
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Ramsay, H.
Direct Submission
Submitted (12-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 15, 2001 this sequence version replaced gi:13751565.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such

```

as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accessions given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TRMBL; Wp., WORMPEP; information on the WORMPEP database can be found at [http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep). This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6>. RP11-321N4 is from the library RPCI-11.2 constructed by the group of Peter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>. VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-321N4. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap. The true left end of clone RP1-3J17 is at 128911 in this sequence. The true right end of clone RP11-30P6 is at 100 in this sequence.

## FEATURES

```

Location/Qualifiers
1..129010
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-321N4"
/clone_lib="RPCI-11.2"
2..192
/feature="MER21B repeat: matches 3..207 of consensus"
repeat_region
240..909
/feature="MER21B repeat: matches 30..715 of consensus"
repeat_region
1593..1682
/feature="MIR repeat: matches 35..145 of consensus"
repeat_region
2030..2102
/feature="L2 repeat: matches 2681..2750 of consensus"
repeat_region
3459..3574
/feature="FLAM_C repeat: matches 1..117 of consensus"
repeat_region
5635..5670
/feature="LIME repeat: matches 683..718 of consensus"
repeat_region
6887..6982
/feature="L2 repeat: matches 2616..2704 of consensus"
repeat_region
8284..8356
/feature="L2 repeat: matches 2639..2710 of consensus"
repeat_region
8380..8405
/feature="L3 copies 2 mer tt 100% conserved"
repeat_region
8440..8601
/feature="MIR repeat: matches 1..169 of consensus"
repeat_region
8825..9360
/feature="L2 repeat: matches 1686..2301 of consensus"
repeat_region
9423..9592
/feature="L2 repeat: matches 2560..2727 of consensus"
repeat_region
12088..12175
/feature="44 copies 2 mer at 73% conserved"
repeat_region
12110..12177
/feature="17 copies 4 mer at 79% conserved"
repeat_region
12790..12905
/feature="L2 repeat: matches 2628..2750 of consensus"
repeat_region
13672..13923
/feature="L2 repeat: matches 2462..2725 of consensus"
repeat_region
14774..15053
/feature="AluSg repeat: matches 1..289 of consensus"
repeat_region
15966..16031
/feature="6 copies 11 mer 77% conserved"
repeat_region
17304..17359
/feature="MIR repeat: matches 196..251 of consensus"
repeat_region
17554..17682
/feature="MIR repeat: matches 108..228 of consensus"
repeat_region
18366..18701
/feature="MER2 repeat: matches 1..345 of consensus"
repeat_region
18939..19248
/feature="AluIo repeat: matches 1..284 of consensus"

```

```
repeat_region 20386..20592
/notes="L2 repeat: matches 70..288 of consensus"
repeat_region 20609..20778
/notes="LTR28 repeat: matches 848..1013 of consensus"
repeat_region 20747..21117
/notes="LTR1 repeat: matches 295..680 of consensus"
repeat_region 21163..21226
/notes="LTR28 repeat: matches 297..361 of consensus"
repeat_region 21187..21359
/notes="LTR1 repeat: matches 16..188 of consensus"
repeat_region 22191..22355
/notes="FRAM repeat: matches 4..163 of consensus"
repeat_region 22844..22991
/notes="MIR repeat: matches 48..202 of consensus"
repeat_region 23155..23190
/notes="18 copies 2 mer ga 86% conserved"
repeat_region 24390..24440
/notes="L2 repeat: matches 2657..2708 of consensus"
repeat_region 25082..25109
/notes="7 copies 4 mer aaac 100% conserved"
repeat_region 25438..25744
/notes="L2 repeat: matches 2427..2744 of consensus"
repeat_region 25751..25917
/notes="MIR repeat: matches 95..262 of consensus"
repeat_region 25964..26307
/notes="L2 repeat: matches 1986..2366 of consensus"
repeat_region 26526..26673
/notes="FRAM C repeat: matches 1..142 of consensus"
repeat_region 26682..26737
/notes="14 copies 4 mer agga 76% conserved"
repeat_region 28208..28394
/notes="MIR repeat: matches 12..196 of consensus"
repeat_region 29052..29250
/notes="MIR repeat: matches 48..251 of consensus"
repeat_region 30082..30170
/notes="MIR repeat: matches 50..145 of consensus"
repeat_region 30197..31090
/notes="MER11D repeat: matches 1..897 of consensus"
repeat_region 31921..32067
/notes="FRAM repeat: matches -2..151 of consensus"
repeat_region 32545..32678
/notes="L2 repeat: matches 2578..2708 of consensus"
repeat_region 32860..32895
/notes="9 copies 4 mer ttta 83% conserved"
repeat_region 32897..33193
/notes="AluX repeat: matches 1..296 of consensus"
repeat_region 35445..35478
/notes="MIR repeat: matches 84..118 of consensus"
repeat_region 37442..37598
/notes="MER5B repeat: matches 1..173 of consensus"
repeat_region 38659..38860
/notes="MIR repeat: matches 15..236 of consensus"
repeat_region 40552..40911
/notes="THE1B repeat: matches 1..364 of consensus"
repeat_region 40912..42540
/notes="THE1B-INTERNAL repeat: matches 1..1580 of consensus"
repeat_region 42541..42899
/notes="THE1B repeat: matches 1..364 of consensus"
repeat_region 42900..43231
/notes="MLT1A repeat: matches 4..349 of consensus"
repeat_region 43407..43851
/notes="MLT1C repeat: matches 9..466 of consensus"
repeat_region 44641..45020
/notes="MST1A repeat: matches 1..382 of consensus"
repeat_region 45689..45988
/notes="AluSg repeat: matches 1..302 of consensus"
repeat_region 46220..46351
/notes="Tiger3b repeat: matches 567..698 of consensus"
repeat_region 46358..46609
/notes="Aluub repeat: matches 40..290 of consensus"
repeat_region 46623..46758
/notes="AluJo/FRAM repeat: matches 170..301 of consensus"
```

```
repeat_region 46759..47229
/notes="Tiger3b repeat: matches 2..540 of consensus"
repeat_region 47230..47442
/notes="MER44A repeat: matches 85..327 of consensus"
repeat_region 47715..48005
/notes="L1PA10 repeat: matches 5864..6153 of consensus"
repeat_region 48577..48973
/notes="WL1B repeat: matches 1..390 of consensus"
repeat_region 50633..50925
/notes="AluY repeat: matches 3..295 of consensus"
repeat_region 53828..55056
/notes="L1MC1 repeat: matches 5023..6332 of consensus"
repeat_region 55092..56081
/notes="L1M4 repeat: matches 3778..4798 of consensus"
repeat_region 56074..56651
/notes="L1MCA repeat: matches 579..1150 of consensus"
repeat_region 56652..56952
/notes="AluX repeat: matches 1..301 of consensus"
repeat_region 56953..57277
/notes="L1MCA repeat: matches 253..579 of consensus"
repeat_region 57280..57562
/notes="L2 repeat: matches 1590..1886 of consensus"
repeat_region 57699..57766
/notes="L2 repeat: matches 2639..2710 of consensus"
repeat_region 59295..59599
/notes="AluX repeat: matches 1..307 of consensus"
repeat_region 59749..59811
/notes="AluX repeat: matches 1..307 of consensus"

Query Match 25.7%; Score 512; DB 9; Length 129010;
Best Local Similarity 100.0%; Pred. No. 14e-244;
Matches 512; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1473 AGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATGTAACACTTGGAATTCGTATAG 1532
Db 50319 AGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATGTAACACTTGGAATTCGTATAG 50260
QY 1533 AATACCCATGAATTTCTGCTGCGAGGCTGTAGAAATTTTACATTTTTCGGGTATAT 1592
Db 50259 AATACCCATGAATTTCTGCTGCGAGGCTGTAGAAATTTTACATTTTTCGGGTATAT 50200
QY 1593 TCTTATATATATATGATCATCGCTGCTGAAATTTTGTAGTTATTTTGTGTTTAAATAAA 1652
Db 50199 TCTTATATATATATGATCATCGCTGCTGAAATTTTGTAGTTATTTTGTGTTTAAATAAA 50140
QY 1653 GACTAACACAACTTAATGATTAAGTAAGTGAATGAGTCTCATAGCTCTTTCATTTCTAGCT 1712
Db 50139 GACTAACACAACTTAATGATTAAGTAAGTGAATGAGTCTCATAGCTCTTTCATTTCTAGCT 50080
QY 1713 GTGATCCAAATTTTATAGAACATAGTCACTTGTGTTATGTCATTTTAAAGAGAAAT 1772
Db 50079 GTGATCCAAATTTTATAGAACATAGTCACTTGTGTTATGTCATTTTAAAGAGAAAT 50020
QY 1773 TCATAATGATCTTATGGCAACAGATAGAGCTGATATAAATTCGTATTGTATAGCTTTGAA 1832
Db 50019 TCATAATGATCTTATGGCAACAGATAGAGCTGATATAAATTCGTATTGTATAGCTTTGAA 49960
QY 1833 AATAATTTATGCTAGTATGAGAAACAGGATAAGATCTGATTTTCTTAGAGTTAATATA 1892
Db 49959 AATAATTTATGCTAGTATGAGAAACAGGATAAGATCTGATTTTCTTAGAGTTAATATA 49900
QY 1893 TTTTAGTAGATGTTGTTTCCCTTTTATTTTATTTTGTACATAGTTAACTGTGTATCTATAA 1952
Db 49899 TTTTAGTAGATGTTGTTTCCCTTTTATTTTATTTTGTACATAGTTAACTGTGTATCTATAA 49840
QY 1953 TAAAGCATCTTATATAGTTTATTAATAATAA 1984
Db 49839 TAAAGCATCTTATATAGTTTATTAATAATAA 49808

RESULT 12
BD113855
LOCUS
DEFINITION EST and encoded human protein.
425 bp
DNA
linear
PAT 18-SEP-2002
```

ACCESSION BD113855  
VERSION BD113855.1 GI:23208759  
KEYWORDS JP 2002010789-A/5932.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 425)  
REFERENCE Edwards, J.B.D.M., Jobert, S. and Giordano, J.E.  
AUTHORS EST and encoded human protein  
TITLE Patent: JP 2002010789-A 5932 15-JAN-2002;  
JOURNAL GENSET CORP  
COMMENT OS Homo sapiens (human)  
PN JP 2002010789-A/5932  
PD 15-JAN-2002  
PF 07-AUG-2000 JP 2000280989  
PR 05-AUG-1999 US 60/147499  
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, SEVELIN JOBERT, JEAN EYE PI  
GIORDANO  
PC C12N15/09, C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, PC  
C12N1/21,  
PC C12N5/10, C12P21/02, C12P21/08, C12Q1/68, C12N15/00, C12N5/00, PC  
C12N15/00  
CC EST and encoded human protein  
FH Key Location/Qualifiers  
FT source 1..425  
FT Location/Qualifiers  
FT source 1..425  
FT /organism="Homo sapiens (human)"  
FT /organism="Homo sapiens"  
FT /mol\_type="genomic DNA"  
FT /db\_xref="taxon:9606"  
BASE COUNT 134 a 82 c 85 g 124 t  
ORIGIN  
Query Match 21.38; Score 425; DB 6; Length 425;  
Best Local Similarity 100.0%; Pred. No. 8.3e-201; Mismatches 0; Indels 0; Gaps 0;  
Matches 425; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1004 CAGAAATATTTATGAGGATGATGACCTGTAGAGGAGTCTATGATTAACCTGATGATGTA 1063  
DB 1 CAGAAATATTTATGAGGATGATGACCTGTAGAGGAGTCTATGATTAACCTGATGATGTA 60  
QY 1064 GGACGGTAGTTTCCAGGTTCCCTGACTGCTATCACTCTTTAATGGGAATCGAATC 1123  
DB 61 GGACGGTAGTTTCCAGGTTCCCTGACTGCTATCACTCTTTAATGGGAATCGAATC 120  
QY 1124 CTCTTTAAACACCCCTGGAATGTATGATTAATCTCTTCACTGATTAATCTGATTAATCTG 1183  
DB 121 CTCTTTAAACACCCCTGGAATGTATGATTAATCTCTTCACTGATTAATCTGATTAATCTG 180  
QY 1184 CTATTTCCAGGAGCACCGTTTGGTCTCACTCATCACTCTTCACTGATGATGATGATGATG 1243  
DB 181 CTATTTCCAGGAGCACCGTTTGGTCTCACTCATCACTCTTCACTGATGATGATGATGATG 240  
QY 1244 GAAACACTGAACTCGCTCTCTCCAGATAAGCAAAAGAGCAAAACAGACTTTTGA 1303  
DB 241 GAAACACTGAACTCGCTCTCTCCAGATAAGCAAAAGAGCAAAACAGACTTTTGA 300  
QY 1304 GAAATGATGATTAATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1363  
DB 301 GAAATGATGATTAATCACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
QY 1364 GAAAGCATCAGACTCTGTTTGGCTTACGACCACTGATCTCAACAGCAGCTGACT 1423  
DB 361 GAAAGCATCAGACTCTGTTTGGCTTACGACCACTGATCTCAACAGCAGCTGACT 420  
QY 1424 TATGT 1428  
DB 421 TATGT 425

## RESULT 13

AX331384/C  
LOCUS AX331384 451 bp DNA linear PAT 09-JAN-2002  
DEFINITION Sequence 1893 from Patent WO0194629.  
ACCESSION AX331384  
VERSION AX331384.1 GI:18122018  
KEYWORDS Homo sapiens (human)  
SOURCE  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,  
AUTHORS Horigan, S., Soppet, D.R. and Weaver, Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
JOURNAL gene sets  
Patent: WO 0194629-A 1893 13-DEC-2001;  
Avalon Pharmaceuticals (US)  
FEATURES Location/Qualifiers  
source 1..451  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
BASE COUNT 179 a 73 c 52 g 146 t 1 others  
ORIGIN  
Query Match 19.38; Score 385; DB 6; Length 451;  
Best Local Similarity 100.0%; Pred. No. 8.3e-181; Mismatches 0; Indels 0; Gaps 0;  
Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1597 ATATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1656  
DB 385 ATATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 326  
QY 1657 AACCAAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1716  
DB 325 AACCAAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 266  
QY 1717 TCCAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1776  
DB 265 TCCAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 206  
QY 1777 AATGATGTTATGCAAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1836  
DB 205 AATGATGTTATGCAAAACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 146  
QY 1837 ATTATGCTAGTATGAGAAACAGAAATAGATGATGATGATGATGATGATGATGATGATGATGAT 1896  
DB 145 ATTATGCTAGTATGAGAAACAGAAATAGATGATGATGATGATGATGATGATGATGATGATGAT 86  
QY 1897 AGTAGATGTTTCCCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 1956  
DB 85 AGTAGATGTTTCCCTTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 26  
QY 1957 GCATCCCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1981  
DB 25 GCATCCCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1  
RESULT 14  
LOCUS BD026500 358 bp DNA linear PAT 27-AUG-2002  
DEFINITION Sequence tag and encoded human protein.  
ACCESSION BD026500  
VERSION BD026500.1 GI:22567723  
KEYWORDS JP 2001269182-A/2746.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens (human)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 358)  
REFERENCE Edwards, J.B.D.M., Duclair, E. and Jordan, J.Y.  
AUTHORS Sequence tag and encoded human protein  
TITLE Patent: JP 2001269182-A 2746 02-OCT-2001;  
JOURNAL

```
GENSET
OS Homo sapiens (human)
PN JP 2001269182-A/2746
PD 02-OCT-2001
PF 24-FEB-2000 JP 2000118773
PI JEAN BAPTISTE DUMAS MILNE EDWARDS, EIMERIC DUCLAIR, JEAN YVES
PC C12N15/09.C07K14/435.C07K16/18.C12N1/15.C12N1/19.C12N1/21, PC
PC C12P21/02.C12P21/08.C12Q1/68//G06F17/30.C12N15/00.C12N5/00, PC
G06F15/40
CC
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
1..358
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 125 a 63 c 69 g 101 t
ORIGIN
Query Match 17.9%; Score 357; DB 6; Length 358;
Best Local Similarity 100.0%; Pred. No. 8.6e-167;
Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 760 ACTACAGATGTAATCTTGGAATAATATAAATCTGTTCTCGGAAACTAATGAAAGA 819
Db 1 ACTACAGATGTAATCTTGGAATAATATAAATCTGTTCTCGGAAACTAATGAAAGA 60
QY 820 GAAAGGTGAGCATTTGGAACCTTTATCATGAATTTCAATTAATCTTGAGTCTCCAA 879
Db 61 GAAAGGTGAGCATTTGGAACCTTTATCATGAATTTCAATTAATCTTGAGTCTCCAA 120
QY 880 GCCTAAACCAAGTAGACAGACATGACCATCTCAGCCCTACTTCAGAAACCAACAGAA 939
Db 121 GCCTAAACCAAGTAGACAGACATGACCATCTCAGCCCTACTTCAGAAACCAACAGAA 180
QY 940 GCTTTTCAATGATCTGTTTAAATAATGCAACCGTGCTGAAATAACAGAGAGAAAGCA 999
Db 181 GCTTTTCAATGATCTGTTTAAATAATGCAACCGTGCTGAAATAACAGAGAGAAAGCA 240
QY 1000 AAATCAGAAATATTTATGAGGTGATGACCTGTAGAGAGAGTCTATGATTTACCTGATGA 1059
Db 241 AAATCAGAAATATTTATGAGGTGATGACCTGTAGAGAGAGTCTATGATTTACCTGATGA 300
QY 1060 TGTAGGACGGTAGTTTCCAGGTTCCGAGTTCCTGAGTTCATCATCTCTTAATGGGAC 1116
Db 301 TGTAGGACGGTAGTTTCCAGGTTCCGAGTTCCTGAGTTCATCATCTCTTAATGGGAC 357

RESULT 15
G30543
LOCUS human STS SHGC-37370, sequence tagged site. 454 bp DNA linear STS 05-OCT-1996
DEFINITION human STS SHGC-37370, sequence tagged site.
ACCESSION G30543
VERSION G30543.1 GI:1594094
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Myers,R.M.
JOURNAL 1 (bases 1 to 454)
Unpublished (1996)
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
```

Email: myers@shgc.stanford.edu

Primer A: ACTGACATCTTCTCAGACTTT

Primer B: GCAATCAATACTCCCAAA

STS size: 150

PCR Profile:

Initial incubation: 94 degrees C for 90 seconds

Denaturation: 94 degrees C for 15 seconds

Annealing: 62 degrees C for 23 seconds

Polymerization: 72 degrees C for 30 seconds

PCR Cycles: 30

Thermal Cycler: Perkin Elmer 9600

Protocol:

Template: 25 ng

Primer: each 1 uM

dNTPs: each 200 uM

Taq Polymerase: 0.05 units/ul

Total Vol: 10 ul

Buffer:

MgCl2: 2.5 mM

KCl: 50 mM

Tris-HCl: 20 mM

pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from N48004  
-- Washington University/Merck EST sequence.

FEATURES  
source

Location/Qualifiers

1..454

/organism="Homo sapiens"

/mol\_type="genomic DNA"

/db\_xref="taxon:9606"

/map="6"

1..150

primer bind

primer\_bind complement(131..150)

BASE COUNT 129 a 79 c 97 g 147 t 2 others

ORIGIN

Query Match

Best Local Similarity 7.2%; Score 143; DB 11; Length 454;

Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1324 AGATCTGTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGT 1383

Db 259 AGATCTGTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGT 318

QY 1384 TGATGGCTTACAGCAACCAAGTACTCAACAAGCAGCTGACTTATGTTTTATTGGACATTGT 1443

Db 319 TGATGGCTTACAGCAACCAAGTACTCAACAAGCAGCTGACTTATGTTTTATTGGACATTGT 378

QY 1444 GATACAGGAACCTGTTTCCAGAGC 1466

Db 379 GATACAGGAACCTGTTTCCAGAGC 401

Search completed: January 31, 2004, 18:20:55

Job time : 7423 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: January 31, 2004, 15:43:54 ; Search time 4299 Seconds  
(without alignments)  
11261.809 Million cell updates/sec

Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: 1 gtafgaaaccccaagtt.....tttaataataaaaaaaaaa 1992

Scoring table: OLIGO NUC  
Gapop 60.0, Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Listing first 45 summaries

Database : EST:

- 1: em\_estba:\*\*
- 2: em\_esthum:\*\*
- 3: em\_estin:\*\*
- 4: em\_estmu:\*\*
- 5: em\_estov:\*\*
- 6: em\_estpl:\*\*
- 7: em\_estro:\*\*
- 8: em\_htc:\*\*
- 9: gb\_est1:\*\*
- 10: gb\_est2:\*\*
- 11: gb\_htc:\*\*
- 12: gb\_est3:\*\*
- 13: gb\_est4:\*\*
- 14: gb\_est5:\*\*
- 15: em\_estfun:\*\*
- 16: em\_estom:\*\*
- 17: em\_ges\_hum:\*\*
- 18: em\_ges\_inv:\*\*
- 19: em\_ges\_pln:\*\*
- 20: em\_ges\_vrt:\*\*
- 21: em\_ges\_fun:\*\*
- 22: em\_ges\_mam:\*\*
- 23: em\_ges\_mus:\*\*
- 24: em\_ges\_pro:\*\*
- 25: em\_ges\_rod:\*\*
- 26: em\_ges\_pbg:\*\*
- 27: em\_ges\_vrl:\*\*
- 28: gb\_ges1:\*\*
- 29: gb\_ges2:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	706	35.4	1096	12	BQ053127
2	690	34.6	702	14	CB054302
3	680	34.1	921	13	BQ230190
4	673	33.8	773	12	BM982264

RESULT 1  
BQ053127  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

BQ053127  
AGENCOURT\_6821786 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5934711  
5', mRNA sequence.  
BQ053127  
BQ053127.1 GI:19812467  
EST.  
Homo sapiens (human)  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1096)  
NIH-MGC <http://mgs.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished  
Contact: Robert Strausberg, Ph.D.  
Email: [cgabs@mail.nih.gov](mailto:cgabs@mail.nih.gov)  
Tissue Procurement: Dr. Daniel McVicar, DBS/NCI  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
<http://image.llnl.gov>  
Plate: L1CM2121 row: e column: 16  
High quality sequence stop: 675.

ALIGNMENTS

652	32.7	664	14	CB136167
651	32.7	917	13	BX372281
640	32.1	708	13	BUE634170
636	31.9	643	14	CB114674
620	31.1	1105	10	BG256631
619	31.1	802	14	CA431124
613	30.8	629	13	BUE579763
608	30.5	608	12	BM829345
606	30.4	606	12	BM699448
599	30.1	604	14	CB130920
577	29.0	594	13	BUE738240
571	28.7	598	13	BUE63958
570	28.6	647	13	BUE679457
560	28.1	986	13	BUE68735
559	28.1	559	12	BM693800
557	28.0	608	12	BM722250
548	27.5	548	14	CB139915
547	27.5	564	12	BM670092
537	27.0	1076	10	BE874550
534	26.8	625	10	BG536529
529	26.6	817	10	BF692296
524	26.3	717	13	BUE59574
513	25.8	513	14	CB142612
510	25.5	558	14	CB160118
506	25.4	506	12	BM835779
505	25.4	616	14	CB054303
503	25.3	503	9	AI283088
503	25.3	503	9	AI955217
497	24.9	1043	9	AL574298
494	24.8	494	10	BG547215
494	24.8	726	10	BG432502
489	24.5	520	9	AW149014
487	24.4	638	10	AW957704
483	24.2	507	14	CA314453
480	24.1	531	9	AI949494
479	24.0	640	9	AW327722
479	24.0	903	9	AL578668
475	23.8	818	10	BG191447
473	23.7	473	14	CB135556
472	23.7	480	9	AI127501
471	23.6	479	10	AW996088

linear EST 29-MAR-2002  
BQ053127  
AGENCOURT\_6821786 NIH\_MGC\_106 Homo sapiens cDNA clone IMAGE:5934711



```
FEATURES
source
Location/Qualifiers
1..1096
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5934711"
/tissue_type="natural killer cells, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 106"
/notes="Organ: blood; Vector: pOTB7; Site 1: XhoI; Site 2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."
BASE COUNT 355 a 226 g 305 t 17 others
ORIGIN
Query Match 35.4%; Score 706; DB 12; Length 1096;
Best Local Similarity 100.0%; Pred. No. 8.9e-235;
Matches 706; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 779 GGGAAAATTATAAATCTGTTCTCGAAAATTAATGAAGAGAAAGTCAGCATTTGGAA 838
DB 10 GGGAAAATTATAAATCTGTTCTCGAAAATTAATGAAGAGAAAGTCAGCATTTGGAA 69
QY 839 CCTTTATCATGAATTTCAATTAATCTTGAGTCTCCAAAGCTTAACCAAGTAGACCA 898
DB 70 CCTTTATCATGAATTTCAATTAATCTTGAGTCTCCAAAGCTTAACCAAGTAGACCA 129
QY 899 GAACGTGACCATCTTCAGCCCTACTTCAGAAAACAAAGAGAGCTTTTCAATGATCTGTTT 958
DB 130 GAACGTGACCATCTTCAGCCCTACTTCAGAAAACAAAGAGAGCTTTTCAATGATCTGTTT 189
QY 959 AAAAATTAATGCAACCGTGTGAAAATACAGAGAGAAAGCAAAATCAGAAATTAATTTATG 1018
DB 190 AAAAATTAATGCAACCGTGTGAAAATACAGAGAGAAAGCAAAATCAGAAATTAATTTATG 249
QY 1019 GAGGTGATGACTGATGAAGAGTCTATGATTAATCTGATGATGATGATGATGATGATGATG 1078
DB 250 GAGGTGATGACTGATGAAGAGTCTATGATTAATCTGATGATGATGATGATGATGATGATG 309
QY 1079 CAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1138
DB 310 CAGGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 369
QY 1139 CTGGAATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1198
DB 370 CTGGAATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 429
QY 1199 CGTTTGGTCTCACTATAACACTTCTCAGAGATGCTATATTTCTGTGAAAACACTGAACCT 1258
DB 430 CGTTTGGTCTCACTATAACACTTCTCAGAGATGCTATATTTCTGTGAAAACACTGAACCT 489
QY 1259 CGCTCTCTCCAGATAGCAAAAGGAGCAAAACAGACTTTTTCAGAAATGATGATTTAC 1318
DB 490 CGCTCTCTCCAGATAGCAAAAGGAGCAAAACAGACTTTTTCAGAAATGATGATTTAC 549
QY 1319 ATTCAGATCTGTTAGTCAAGTGTATTTGTTGAAGAAACCAAGTATGAAGCATCAGACTT 1378
DB 550 ATTCAGATCTGTTAGTCAAGTGTATTTGTTGAAGAAACCAAGTATGAAGCATCAGACTT 609
QY 1379 CTGTTTGAATGGCTTTCAGCAACCAAGTACTCAACAGCAGCTGACTTATGTTTATTTGGAC 1438
DB 610 CTGTTTGAATGGCTTTCAGCAACCAAGTACTCAACAGCAGCTGACTTATGTTTATTTGGAC 669
QY 1439 ATTGTGATACAGGAACCTGTTTCCAGAGCTCAATAGGTACAAAAG 1484
DB 670 ATTGTGATACAGGAACCTGTTTCCAGAGCTCAATAGGTACAAAAG 715
```

---

```
RESULT 2
CB054302/c
LOCUS
DEFINITION NISC_gm03e09.x1 NCI_CGAP_Brn23 Homo sapiens cDNA clone
IMAGE:3291376 3', mRNA sequence.
ACCESSION CB054302
VERSION CB054302.1 GI:27792589
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 702)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
cDNA Library Preparation:
cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC)
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
info@image.llnl.gov
Plate: LLAM8060 row: J column: 17
Seq primer: -21M13 forward primer (ABI).
FEATURES
source
Location/Qualifiers
1..702
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3291376"
/tissue_type="glioblastoma (pooled)"
/lab_host="DH10B"
/clone_lib="NCI_CGAP_Brn23"
/notes="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGGCGCGCATATCTTTTGTGTTTGTGTTTGTGTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73-vector.
Library is normalized, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 253 a 121 c 86 g 242 t
ORIGIN
Query Match 34.6%; Score 690; DB 14; Length 702;
Best Local Similarity 100.0%; Pred. No. 4.2e-229;
Matches 690; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1303 AGAATGATGAATTAATTCAGATCTGTTAGTCAAGTGTATGTTGTAAGAACCAAGTA 1362
DB 702 AGAATGATGAATTAATTCAGATCTGTTAGTCAAGTGTATGTTGTAAGAACCAAGTA 643
QY 1363 TGAAGCATCAGACTTCTGTTTGTGTTGCTTACACACCACTACTCAACAGCAGCTGAC 1422
DB 642 TGAAGCATCAGACTTCTGTTTGTGTTGCTTACACACCACTACTCAACAGCAGCTGAC 583
QY 1423 TTATGTTTATTTGACATTTGTGATACAGGAATCTTTCCAGAGCTCAATAGGTACAAA 1482
DB 582 TTATGTTTATTTGACATTTGTGATACAGGAATCTTTCCAGAGCTCAATAGGTACAAA 523
QY 1483 GGAAGTTACCTCTGTGACATCTTGATGTAAACACTTGGATTTGGTATAGAAATACCAT 1542
DB 522 GGAAGTTACCTCTGTGACATCTTGATGTAAACACTTGGATTTGGTATAGAAATACCAT 463
QY 1543 TGAATTTCTCTGTGGAGGCTGTAGAAATTTACTTTTTTGGGTATTTCTTATATAT 1602
DB 462 TGAATTTCTCTGTGGAGGCTGTAGAAATTTACTTTTTTGGGTATTTCTTATATAT 403
```



University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clones Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com) or from Open Biosystems  
(www.openbiosystems.com).  
Seq primer: M13 FORWARD  
POLYA-Yes.

FEATURES  
source  
1. .773  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-CF-EN1-acr-j-04-0-UI"  
/tissue\_type="Primary Lung Cystic Fibrosis Epithelial  
Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-CF-EN1"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
modified polylinker; Site 1: EcoRI; Site 2: Not I;  
UI-CF-EN1 is a normalized cDNA library containing the  
following tissue(s): Primary Lung Cystic Fibrosis  
Epithelial Cells. The library was constructed according to  
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,  
1996. First strand cDNA synthesis was primed with an  
oligo-dT primer containing a Not I site. Double stranded  
cDNA was ligated to an EcoRI adaptor, digested with Not  
I, and cloned directionally into pT73-Pac vector. The  
oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is  
located between the Not I site and the (dT)18 tail. The  
sequence tag for this library is CTGCTCAGGT.  
TAG LIB=UI-CF-EN1  
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS  
6hr to LPS 24h  
TAG\_SEQ=CTGCTCAGGT  
233 a 146 g 269 t 1 others

Query Match 33.8%; Score 673; DB 12; Length 773;  
Best Local Similarity 100.0%; Pred. No 3.1e-223; Mismatches 0; Indels 0; Gaps 0;  
Matches 673; Conservative 0;  
Qy 1010 TATTTATGAGGATGACTGTAGAGGAGTCTATGATTAACCTGATGTATGAGACGG 1069  
Db 690 TATTTATGAGGATGACTGTAGAGGAGTCTATGATTAACCTGATGTATGAGACGG 631  
Qy 1070 GTAGTTTCCAGGTTCTGACTGCTTCATCTCTTAATGGGAACCTGAATCTCTTT 1129  
Db 630 GTAGTTTCCAGGTTCTGACTGCTTCATCTCTTAATGGGAACCTGAATCTCTTT 571  
Qy 1130 AAAAAACCTCGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1189  
Db 570 AAAAAACCTCGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 511  
Qy 1190 CAGGAGCAGCGTTGGTCTCACTCATACACTTCTCAGAGATGATGATGATGATGAT 1249  
Db 510 CAGGAGCAGCGTTGGTCTCACTCATACACTTCTCAGAGATGATGATGATGATGAT 451  
Qy 1250 ACTGAACCTCGCTCTCTCCAGATTAAGCAAAAAAGGAGCAAAACAGACTTTTGAAGAAATG 1309  
Db 450 ACTGAACCTCGCTCTCTCCAGATTAAGCAAAAAAGGAGCAAAACAGACTTTTGAAGAAATG 391  
Qy 1310 ATGAATTACATCCAGATCTGTAGTCAAGTATGATGATGATGATGATGATGATGAT 1369  
Db 390 ATGAATTACATCCAGATCTGTAGTCAAGTATGATGATGATGATGATGATGATGAT 331

Qy 1370 ATCAGACTCTCTGTTTGTATGCTTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTT 1429  
Db 330 ATCAGACTCTCTGTTTGTATGCTTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTT 271  
Qy 1430 TTATTTGACATGTTGATACAGCAACAGTACTCAACAGCAGCTGACTTATGTT 1489  
Db 270 TTATTTGACATGTTGATACAGCAACAGTACTCAACAGCAGCTGACTTATGTT 211  
Qy 1490 ACCTCTGTGACATCTTGGATGTAAACACATGTTGGATTTGGATGTAAACACATGTTAAAT 1549  
Db 210 ACCTCTGTGACATCTTGGATGTAAACACATGTTGGATTTGGATGTAAACACATGTTAAAT 151  
Qy 1550 TCTGCTGTGAGGAGTGTAGAAATTTACATTTTGGTGTATATTTCTTATATATATATGTT 1609  
Db 150 TCTGCTGTGAGGAGTGTAGAAATTTACATTTTGGTGTATATTTCTTATATATATATGTT 91  
Qy 1610 ACATCGCTGTCTGAATTTTAGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTA 1669  
Db 90 ACATCGCTGTCTGAATTTTAGTATTTTGTATTTTGTATTTTGTATTTTGTATTTTGTATTTA 31  
Qy 1670 TGATTAAGTGA 1692  
Db 30 TGATTAAGTGA 18

RESULT 5  
CB136167  
LOCUS  
DEFINITION X-EST0188577 L14ChoICKO Homo sapiens cDNA clone L14ChoICKO-20-H05  
5' mRNA sequence.  
ACCESSION CB136167  
VERSION CB136167.1 GI:28103366  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 664)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 20 row: H column: 05  
High quality sequence stop: 664.  
Location/Qualifiers  
1. .664  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="L14ChoICKO-20-H05"  
/sex="M"  
/cell\_line="Choi-CK"  
/lab\_host="Top10F"  
/clone\_lib="L14ChoICKO"  
/note="Organ: Liver; Vector: pT73-Pac; Site 1: EcoRI;  
Site 2: NotI; The library was contributed by the Soares  
laboratory and it was constructed as described by Bonaldi,  
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research  
6(9): 791-806. RNA was prepared from harvested cell  
culture."

FEATURES  
source

BASE COUNT 227 a 107 c 146 g 184 t

ORIGIN

Query Match 32.7%; Score 652; DB 14; Length 664;



RESULT	7
BU634170/c	
LOCUS	
DEFINITION	708 bp mRNA linear EST 23-SEP-2002
ACCESSION	BU634170
KEYWORDS	UI-H-FLI-bgw-1-08-0-UI.s1 NCI CGAP FLI Homo sapiens cDNA clone
SOURCE	UI-H-FLI-bgw-1-08-0-UI 3', mRNA sequence.
ORGANISM	BU634170.1 GI:23301425
REFERENCE	EST.
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
JOURNAL	Tumor Gene Index
COMMENT	Unpublished Contact: Robert Strausberg, Ph.D. Email: cgaps-remail.nih.gov Tissue Procurement: James Martin CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa DNA sequencing by: Dr. M. Bento Soares, University of Iowa Clone distribution: Clone distribution information can be obtained from Dr. M. Bento Soares, bento-soares@uiowa.edu Seq primer: M13 FORWARD POLYA=Yes.
FEATURES	Location/Qualifiers
source	1..708
	/organism="Homo sapiens"
	/mol_type="mRNA"
	/db_xref="taxon:9606"
	/clone="UI-H-FLI-bgw-1-08-0-UI"
	/tissue_types="Cell lines"
	/dev_stage="Adult"
	/lab_host="DR10B (Life Technologies)"
	/clone_lib="NCI CGAP FLI"
	/notes="Organ: Chondrosarcoma; Vector: pT73-Pac (Pharmacia)
	with a modified polylinker; Site 1: EcoR I; Site 2: Not
	I; NCI CGAP FLI is a normalized cDNA library derived from
	a pool of mRNA obtained from 4 cell lines from grade III
	chondrosarcoma tissues. The library was constructed
	according to Bonaldo, Lennon and Soares, Genome Research,
	6:791-806, 1996. First strand cDNA synthesis was primed
	with an oligo-dT primer containing a Not I site. Double
	stranded cDNA was ligated to an EcoR I adaptor, digested
	with Not I, and cloned directionally into pT73-Pac
	vector. The oligonucleotide used to prime the synthesis of
	first-strand cDNA contains a library tag sequence that is
	located between the Not I site and the (dT)18 tail. The
	sequence tag for this library is GAGTCGGTG. The cell
	lines were provided by Dr. James Martin from the
	University of Iowa.
	TAG LIB=UI-H-FLI
	TAG TISSUE=Human Chondrosarcoma Grade 3 cell line mix
	TAG SEQ=GAGGTCGGTG"
BASE COUNT	219 a 135 c 114 g 239 t 1 others
ORIGIN	
Query Match	32.1%; Score 640; DB 13; Length 708;
Best Local Similarity	99.9%; Pred. No. 9.2e-212;
Matches 690; Conservative	0; Mismatches 1; Indels 0; Gaps 0;
QY	992 AGAAGCAAAATCAGATTATTTATGGAGGTGATGACTGTAGAGAGCTATGATTAC 1051
Db	708 AGAAGCAAAATCAGATTATTTATGGAGGTGATGACTGTAGAGAGCTATGATTAC 649
QY	1052 CTGATGATGTAGACGGGTAGTTTCCAGGTCCTCGACTGGCTTCATCTCTTAATG 1111
Db	648 CTGATGATGTAGACGGGTAGTTTCCAGGTCCTCGACTGGCTTCATCTCTTAATG 589
QY	1112 GGAACCTCGAATCCTCTTTTAAAAACACCCCTGGAAATGTATATCTAGTACTATCTTCAGTGT 1171

```

/cell_line="Cho-Ck"
/lab_host="Top10P"
/clone_lib="L10chock0"
/notes="Organ: Liver; Vector: pVT3-Pac; Site 1: EcoRI;
Site 2: NotI; The library was contributed by the Soares
laboratory and it was constructed as described by Bonaldo,
M.F., Lennon, G. and Soares, M.B. (1996), Genome Research
6(9): 791-806. RNA was prepared from harvested cell
culture."
BASE COUNT      214 a   123 c   126 g   180 t
ORIGIN

Query Match      31.9%; Score 636; DB 14; Length 643;
Best Local Similarity 100.0%; Pred. No. 2.4e-210;
Matches 636; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      894  GACCAGAACTGACCACTCTCAGCCCTACTCTCAGAAACAAACAGAGAGCTTTTCAATGATC 953
DB      8    GACCAGAACTGACCACTCTCAGCCCTACTCTCAGAAACAAACAGAGAGCTTTTCAATGATC 67

QY      954  TGTTTAAAAAATGAATGAAACCGTGTGAAATAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1013
DB      68  TGTTTAAAAAATGAATGAAACCGTGTGAAATAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 127

QY      1014 TATGAGAGTGATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1073
DB      128 TATGAGAGTGATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 187

QY      1074 TTTTCCAGGTTCTGAGTCTGCTCATCATCTCTTAATGGAAGTCTGAAATCCCTTTTAAAA 1133
DB      188 TTTTCCAGGTTCTGAGTCTGCTCATCATCTCTTAATGGAAGTCTGAAATCCCTTTTAAAA 247

QY      1134 ACACCTCGGAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1193
DB      248 ACACCTCGGAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 307

QY      1194 AGCAGCGTTGGTCTCAGTCATCAACACTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1253
DB      308 AGCAGCGTTGGTCTCAGTCATCAACACTCTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 367

QY      1254 AACCTGCTCTCTCCAGATAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1313
DB      368 AACCTGCTCTCTCCAGATAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 427

QY      1314 ATTACATTCAGATCTGTTAGTCAAGTGATGTTGGTCAAGAAACCAAGTATGAAGAGATCA 1373
DB      428 ATTACATTCAGATCTGTTAGTCAAGTGATGTTGGTCAAGAAACCAAGTATGAAGAGATCA 487

QY      1374 GACTTCTGTTGATGGCTTACAGCAACCAAGTACTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1433
DB      488 GACTTCTGTTGATGGCTTACAGCAACCAAGTACTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG 547

QY      1434 TGGACATTTGATACAGGAACTGTTTCCAGAGCTCAATAGGTAAGGTAAGGTAAGGTAAGGTAAGGTA 1493
DB      548 TGGACATTTGATACAGGAACTGTTTCCAGAGCTCAATAGGTAAGGTAAGGTAAGGTAAGGTAAGGTA 607

QY      1494 CTGTGACATCTCGAGTAAACACTTGGATTGGTA 1529
DB      608 CTGTGACATCTCGAGTAAACACTTGGATTGGTA 643

```

```

RESULT 9
LOCUS      BG256631
DEFINITION 602370809F1 NIH_MGC_92 Homo sapiens cDNA clone IMAGE:4478858 5',
            mRNA sequence.
ACCESSION  BG256631
VERSION     BG256631.1 GI:12766447
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

# REFERENCE

1 (bases 1 to 1105)  
 NIH-MGC <http://mgs.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM10310 Row: a Column: 03  
 High quality sequence stop: 645.

## FEATURES

### source

1..1105  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4478858"  
 /tissue\_type="embryonal carcinoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /clone\_lib="NIH\_MGC\_92"  
 /notes="Organ: testis; Vector: pCMV-SPORT6; Site 1: NotI;  
 Site 2: SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 2.5 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH MGC Library."  
 BASE COUNT 380 a 235 c 208 g 282 t

### ORIGIN

```

Query Match      31.1%; Score 620; DB 10; Length 1105;
Best Local Similarity 100.0%; Pred. No. 6e-205;
Matches 620; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      764  CCAGATGTAATCTGGGAAAAATATAAAATCTGTCCTGGAAAACTAATGAAGAGAA 823
DB      1    CCAGATGTAATCTGGGAAAAATATAAAATCTGTCCTGGAAAACTAATGAAGAGAA 60

QY      824  GGTGAGATTTGGAACCTTTTATCAGTAATTCATTAATTTCTGTGAGTCTCCAGAGCCT 883
DB      61  GGTGAGATTTGGAACCTTTTATCAGTAATTTCAATTAATTTCTGTGAGTCTCCAGAGCCT 120

QY      884  AAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTCAGAAAAACAAACAGAGAGCTT 943
DB      121 AAACCAAGTAGACAGAACTGACCAATCTCAGCCCTACTTCAGAAAAACAAACAGAGAGCTT 180

QY      944  TTCAATGATCTGTTTAAAAATAATGCAACCGTGTGAAAAATACAGAGAGAGAGAGAGAGAG 1003
DB      181 TTCAATGATCTGTTTAAAAATAATGCAACCGTGTGAAAAATACAGAGAGAGAGAGAGAGAG 240

QY      1004 CAGAAATTTATGAGAGGTGATGACTGTAGAGAGGTCTATGATTACCTGATGATGTA 1063
DB      241 CAGAAATTTATGAGAGGTGATGACTGTAGAGAGGTCTATGATTACCTGATGATGTA 300

QY      1064 GGACGGGTAGTTTCCAGAGTTTCCAGCTGCTCATCATCTCTTAATGGAAGTCTCGAATC 1123
DB      301 GGACGGGTAGTTTCCAGAGTTTCCAGCTGCTCATCATCTCTTAATGGAAGTCTCGAATC 360

QY      1124 CTCTTTAAAAACACCTCGGAATGTATATGATTACTATCTTCAGTGTAACTAGAACAG 1183
DB      361 CTCTTTAAAAACACCTCGGAATGTATATGATTACTATCTTCAGTGTAACTAGAACAG 420

QY      1184 CTATTTGAGGAGCAACCGTTGGTCTCACTCATACACTTCTCAGAGATGCTATATCTGT 1243
DB      421 CTATTTGAGGAGCAACCGTTGGTCTCACTCATACACTTCTCAGAGATGCTATATCTGT 480

QY      1244 GAAACACTGAACTCGCTCTCTCCAGATAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1303
DB      481 GAAACACTGAACTCGCTCTCTCCAGATAGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540

QY      1304 GAAATGATGAATTACATTCCAGATCTGTAGTCAAGTGTATTGGTGAAGAAACCAAGTAT 1363

```



```

Db      541 GAAATGATGATACATTCAGATCTGTAGTCAAGTGTATGGTGAGAGAAACCAAGTAT 600
Qy      1364 GAAAGCATCAGACTTCTGTT 1383
Db      601 GAAAGCATCAGACTTCTGTT 620

RESULT 10
CA431124/c
LOCUS
DEFINITION
  UI-H-FLI-bge-f-14-0-UI.s1 NCI CGAP FLI Homo sapiens cDNA clone
  UI-H-FLI-bge-f-14-0-UI 3', mRNA sequence.
ACCESSION
  CA431124
VERSION
  CA431124.1 GI:24793850
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 802)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: James Martin
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be obtained
from Dr. M. Bento Soares, bento-soares@uiowa.edu
The following repetitive elements were found in this cDNA
sequence: 1-70, >AT richLow_complexity (matched complement)
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES
    source
    1..802
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="UI-H-FLI-bge-f-14-0-UI"
    /tissue_type="Cell lines"
    /dev_stage="Adult"
    /lab_hosts="DH10B (Life Technologies)"
    /clone_lib="NCI CGAP FLI"
    /note="Organ: Chondrosarcoma; Vector: p773-Pac (Pharmacia
    I; NCI CGAP FLI is a normalized cDNA library derived from
    a pool of mRNA obtained from 4 cell lines from grade III
    chondrosarcoma tissues. The library was constructed
    according to Bonaldo, Lennon and Soares, Genome Research,
    6:791-806, 1996. First strand cDNA synthesis was primed
    with an oligo-dT primer containing a Not I site. Double
    stranded cDNA was ligated to an EcoR I adaptor, digested
    with Not I, and cloned directionally into p773-Pac
    vector. The oligonucleotide used to prime the synthesis of
    first-strand cDNA contains a library tag sequence that is
    located between the Not I site and the (dT)18 tail. The
    sequence tag for this library is GAGTCGGTG. The cell
    lines were provided by Dr. James Martin from the
    University of Iowa.
    TAG_LIB=UI-H-FLI
    TAG_TISSUE=Human Chondrosarcoma Grade 3 cell line mix
    TAG_SEQ=GAGTCGGTG"
BASE COUNT 238 a 151 c 136 g 275 t 2 others
ORIGIN

Query Match 31.1%; Score 619; DB 14; Length 802;
Best Local Similarity 99.7%; Pred. No. 1.6e-204;
Matches 719; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      958 TAAAAATATGCAACCGTCTGAAAAATACAGAGAGAAAGCAAAATCAGATTTATTTAT 1017
Db      735 TAAAAATATGCAACCGTCTGAAAAATACAGAGAGAAAGCAAAATCAGATTTATTTAT 676
Qy      1018 GGAGGTGATGACTGTAGAGGAGTCTATGATTACCTGATGTATGTAGGACGGTAGTTT 1077
Db      675 GGAGGTGATGACTGTAGAGGAGTCTATGATTACCTGATGTATGTAGGACGGTAGTTT 616
Qy      1078 CCAGGTTCTGACTGGCTTCATCTCTTAATGGGAACTCGAATCTCTTTAAAAACAC 1137
Db      615 CCAGGTTCTGACTGGCTTCATCTCTTAATGGGAACTCGAATCTCTTTAAAAACAC 556
Qy      1138 CCTGGAAATGTATACTGATTACTATCTTCAGTCTAAACTAGAACAGCTATTTCAGAGCA 1197
Db      555 CCTGGAAATGTATACTGATTACTATCTTCAGTCTAAACTAGAACAGCTATTTCAGAGCA 496
Qy      1198 CCGTTTGGTCTCACTCATACACCTTCTCAGAGATGCTATATTCTGTGAAACACCTGAACC 1257
Db      495 CCGTTTGGTCTCACTCATACACCTTCTCAGAGATGCTATATTCTGTGAAACACCTGAACC 436
Qy      1258 TCGTCTCTCCAGATAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAATTA 1317
Db      435 TCGTCTCTCCAGATAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGAATTA 376
Qy      1318 CATTCCAGATCTGTAGTCAAGTGTATTGGTGAAGAACCAAGTATGAAAGCATCAGACT 1377
Db      375 CATTCCAGATCTGTAGTCAAGTGTATTGGTGAAGAACCAAGTATGAAAGCATCAGACT 316
Qy      1378 TCTGTTTGGTCTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTTTATTTGGA 1437
Db      315 TCTGTTTGGTCTTACAGCAACAGTACTCAACAGCAGCTGACTTATGTTTATTTGGA 256
Qy      1438 CATTGTGATCAGGAATCTGTTCCAGAGCTCAATAAGGTTACAAAGGAAGTACCTCTGT 1497
Db      255 CATTGTGATCAGGAATCTGTTCCAGAGCTCAATAAGTACAAAGGAAGTACCTCTGT 196
Qy      1498 GACATCTGGATGTAAACACCTGGATTGTGTATAGATACCCATGGAATTTCTGCTGT 1557
Db      195 GACATCTGGATGTAAACACCTGGATTGTGTATAGATACCCATGGAATTTCTGCTGT 136
Qy      1558 GCGAGGTGGTGTAGAAATTTACTTTTTGGGTATATTCTTATATATATATATGATCATCGCT 1617
Db      135 GCGAGGTGGTGTAGAAATTTACTTTTTGGGTATATTCTTATATATATATATGATCATCGCT 76
Qy      1618 GTCTGAAATTTAGTATTTTGTGTTTTTAATAGACTAACACAACTTAATGATTA 1677
Db      75 GTCTGAAATTTAGTATTTTGTGTTTTTAATAGACTAACACAACTTAATGATTA 16
Qy      1678 A 1678
Db      15 A 15

RESULT 11
BUS79763/c
LOCUS
DEFINITION
  im90e05.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:612216
  3' similar to TR:Q9Y5W7 Q9Y5W7 SORTING NEXIN 14 ; mRNA sequence.
ACCESSION
  BUS79763
VERSION
  BUS79763.1 GI:23064990
KEYWORDS
  EST.
SOURCE
  Homo sapiens (human)
ORGANISM
  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 629)
REFERENCE
  1
AUTHORS
  Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
  Lemishka,I., Searce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
  Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
  Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas
  ,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.
  , Jackson,Y. and Bowers,Y.
  Endocrine Pancreas Consortium
TITLE

```

JOURNAL  
COMMENT

Unpublished  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-9557  
Email: dmelton@bichp.harvard.edu  
Library was constructed by Dr. J. Ferrer in vivo mass-excised to  
pBluescript SK- by Dr. H. Inoue DNA sequencing by: Washington  
University Genome Sequencing Center For information on obtaining a  
clone please contact: Dr. Hiroshi Inoue (hinoue@im.wustl.edu)  
Seq primer: -40UP from Gibco  
High quality sequence stop: 456.

## FEATURES

Location/Qualifiers  
source  
1..629  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="IMAGE:612216"  
/issue\_type="insulinoma"  
/lab\_host="DH10B (phage-resistant)"  
/clone\_lib="Human insulinoma"  
/note="Organ: pancreas; Vector: pBluescript SK-; Site: 1:  
XhoI; Site 2: EcoRI; Constructed with lambda ZAPII system  
(Stratagene) by Dr. J. Ferrer, in vivo mass-excised to  
pBluescript SK- by Dr. H. Inoue following the Washington  
University protocol  
(http://genome.wustl.edu/est/lambda\_protocol.shtml).  
Please contact Hiroshi Inoue, MD/PhD for further  
information on this library (Metabolism Division, Permutt  
Laboratory, Washington University School of Medicine, Box  
8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this  
is a Washington University Pancreas EST project library."  
BASE COUNT 198 a 113 c 107 g 211 t  
ORIGIN

Query Match 30.8%; Score 613; DB 13; Length 629;  
Best Local Similarity 100.0%; Pred. No. 2.3e-202;  
Matches 613; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1070 GTAGTTTCCAGGTTCTGACTGGCTTCATCATCTTTAATGGAACTCGAATCTCTTT 1129  
DB 629 GTAGTTTCCAGGTTCTGACTGGCTTCATCATCTTTAATGGAACTCGAATCTCTTT 570  
QY 1130 AAAAACCCTCGAATGATATATCTGATTAATCTTCTGATTAATGAACTAGAACTATTT 1189  
DB 569 AAAAACCCTCGAATGATATATCTGATTAATCTTCTGATTAATGAACTAGAACTATTT 510  
QY 1190 CAGGACCCGTTTGGTCTCCTCATAACACTTCTCAGAGATGCTATATCTGTGAAC 1249  
DB 509 CAGGACCCGTTTGGTCTCCTCATAACACTTCTCAGAGATGCTATATCTGTGAAC 450  
QY 1250 ACTGAACCTCGCTCTCTCCAGATAAGCAAAAAGGAGCAAAACAGACTTTTGAAGAAATG 1309  
DB 449 ACTGAACCTCGCTCTCTCCAGATAAGCAAAAAGGAGCAAAACAGACTTTTGAAGAAATG 390  
QY 1310 ATGAATTAATCCAGATCTGTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGC 1369  
DB 389 ATGAATTAATCCAGATCTGTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAGC 330  
QY 1370 ATCAGATCTCTGTTGATGCTTACAGCAACCAAGTACTCAACAGCACTGACTTATGTT 1429  
DB 329 ATCAGATCTCTGTTGATGCTTACAGCAACCAAGTACTCAACAGCACTGACTTATGTT 270  
QY 1430 TTATTGACATTTGTGATACAGGAACTGTTTCCAGAGCTCAATAAGGTACAAAAGGAAGTT 1489  
DB 269 TTATTGACATTTGTGATACAGGAACTGTTTCCAGAGCTCAATAAGGTACAAAAGGAAGTT 210  
QY 1490 ACCTCTGTGACATCTTGGATGTAACCACTTGGATTTGGTATAGAAATPAAACCCATGAAAT 1549  
DB 209 ACCTCTGTGACATCTTGGATGTAACCACTTGGATTTGGTATAGAAATPAAACCCATGAAAT 150

QY 1550 TCTGCTGCTGAGGGTGGTAGAATTTACTTTTTGGGTATATTTCTATATATATATATGT 1609  
DB 149 TCTGCTGCTGAGGGTGGTAGAATTTACTTTTTGGGTATATTTCTATATATATATATGT 90  
QY 1610 ACATCGCTGCTGCTGAAATTTAGTTATTTTGTATTAATAAAGACTAAACAAACTTAA 1669  
DB 89 ACATCGCTGCTGCTGAAATTTAGTTATTTTGTATTAATAAAGACTAAACAAACTTAA 30  
QY 1670 TGATTAAGTGA 1682  
DB 29 TGATTAAGTGA 17

## RESULT 12

BM829345  
LOCUS K-EST0102319 S9SNU601 Homo sapiens cDNA clone S9SNU601-45-A08 5',  
DEFINITION mRNA sequence.  
ACCESSION BM829345  
VERSION BM829345.1 GI:19185754  
KEYWORDS EST.  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 608)  
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished  
COMMENT Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eosun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 45' row: A column: 08  
High quality sequence stop: 608.  
FEATURES  
Location/Qualifiers  
1..608  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="S9SNU601-45-A08"  
/sex="M"  
/tissue\_type="Ascites"  
/cell\_line="Epithelial"  
/lab\_host="Top10P"  
/clone\_lib="S9SNU601"  
/note="Organ: Stomach; Vector: pME18-FL3, Site 1: XhoI;  
Site 2: XhoI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tabacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including SfiI  
site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized with Superscript II using SfiI  
oligo-dT primer. After first strand synthesis, RNA was  
degraded by NaOH treatment and cDNA was amplified by PCR  
reaction. The PCR products were digested with SfiI and  
cloned into DralII- digested pME18-FL3 vector. The  
obtained cDNA vectors were used for transfection of  
competent cells E. coli Top10P by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."

## BASE COUNT

203 a 100 c 130 g 175 t

## ORIGIN

Query Match 30.5%; Score 608; DB 12; Length 608;  
Best Local Similarity 100.0%; Pred. No. 1.3e-200;



Db 481 ACAGCAACAGTACTCAACAGAGCTGACTTATGTTTATTTGGACATTTGATACAGA 540  
 1453 ACTGTTTCCAGAGCTCAATAAGTACAAAGGAAGTACCTCTGTGACATCTTGGATGTA 1512  
 Db 541 ACTGTTTCCAGAGCTCAATAAGTACAAAGGAAGTACCTCTGTGACATCTTGGATGTA 600  
 QY 1513 AACACT 1518  
 Db 601 AACACT 606

RESULT 14  
 CBI30920  
 LOCUS K-EST0180911 L12JSHCO Homo sapiens cDNA clone L12JSHCO-5-A04 5',  
 DEFINITION mRNA sequence.  
 ACCESSION CBI30920  
 VERSION CBI30920.1 GI:28095369  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 604)  
 AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
 Oh,K.J., Cheong,J.S., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
 Kim,Y.S.  
 TITLE 21C Frontier Korean EST Project 2001  
 JOURNAL Unpublished  
 COMMENT Contact: Kim YS  
 Genome Research Center  
 Korea Research Institute of Bioscience & Biotechnology  
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
 Tel: +82-42-860-4470  
 Fax: +82-42-860-4409  
 Email: yongseungemail.kr@kribb.re.kr  
 Plate: 5 row: A column: 04  
 High quality sequence stop: 604.

FEATURES  
 source  
 1..604  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="L12JSHCO-5-A04"  
 /sex="M"  
 /cell\_line="J-SHC"  
 /lab\_host="Top10"  
 /clone\_lib="L12JSHCO"  
 /note="Organ: Liver; Vector: pcNS-D2; Site 1: EcoRI;  
 Site 2: NotI; The poly (A) + RNA was dephosphorylated with  
 bacterial alkaline phosphatase (BAP) and then decapped  
 with tabacco acid pyrophosphatase (TAP). The decapped  
 intact mRNA was ligated with DNA-RNA linker including  
 EcoRI site by treatment of T4 RNA ligase and the first  
 strand cDNA was synthesized from oligo dr-selected mRNA by  
 priming with dr-tailed vector. The dr-tailed vector was  
 adjusted to have about 50nt. The cDNA vector was  
 circularized with E. coli DNA ligase after digestion of  
 EcoRI which site is also included in vector. An RNA strand  
 converted to a DNA strand by Okayama-Berg method. The  
 obtained cDNA vectors were used for transfection of  
 competent cells E. coli Top10' by electroporation method.  
 The cDNA libraries constructed by this method are  
 full-length enriched cDNA library."  
 BASE COUNT 205 a 100 c 137 g 162 t  
 ORIGIN

Query Match 30.1%; Score 599; DB 14; Length 604;  
 Best Local Similarity 100.0%; Pred. No. 1.7e-197;  
 Matches 599; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 145 CAGGACACACAGAAAGGGGAGATCATTTGGATCAGCAGATAGGTAGCAAAATTA 204

Db 6 CAGGACACACAGAAAGGGGAGATCATTTGGATCAGCAGATAGGTAGCAAAATTA 65  
 QY 205 AGGAGTATTTCAAAGTACCAATGAGGGAGCTATGTTGCCTAAATATGCTAGCTGA 264  
 Db 66 AGGAGTATTTCAAAGTACCAATGAGGGAGCTATGTTGCCTAAATATGCTAGCTGA 125  
 QY 265 AGGTGAAGATGATTTTATTAAGAAGGTATTTGTTGAATGAAGATGATTTCTCCAGTGA 324  
 Db 126 AGGTGAAGATGATTTTATTAAGAAGGTATTTGTTGAATGAAGATGATTTCTCCAGTGA 185  
 QY 325 GGCTGTGAGCACACCTAATACTCCCGAAACCTTGTGTCATGAAAATTAAGCATTCATA 384  
 Db 186 GGCTGTGAGCACACCTAATACTCCCGAAACCTTGTGTCATGAAAATTAAGCATTCATA 245  
 QY 385 TGTAGACTTTTTCAGAGATCCCTCTCTGAAGAGGAGGAGAAAAGAAAGAAATTCCTGT 444  
 Db 246 TGTAGACTTTTTCAGAGATCCCTCTCTGAAGAGGAGGAGAAAAGAAAGAAATTCCTGT 305  
 QY 445 GTTTTGTATTGATGTTCAAGAAATGATAGAGAGCTTTGGACACAGAGCTCGAATTCG 504  
 Db 306 GTTTTGTATTGATGTTCAAGAAATGATAGAGAGCTTTGGACACAGAGCTCGAATTCG 365  
 QY 505 GTCTGTCTATAGAGATATCTTGAATTTCTATGTACTTGAATCAAACTAAACAGAAATTC 564  
 Db 366 GTCTGTCTATAGAGATATCTTGAATTTCTATGTACTTGAATCAAACTAAACAGAAATTC 425  
 QY 565 TGGTGCATTTCTGATGCCAGCTTCTCTTAAGAGATCATTTGGCCCCCAAAAATTAATGA 624  
 Db 426 TGGTGCATTTCTGATGCCAGCTTCTCTTAAGAGATCATTTGGCCCCCAAAAATTAATGA 485  
 QY 625 ATTCTTAAGTCAAGAGAGGAGATTTCCAAAGATATCTACAGAACTTCTGCGAGCATCC 684  
 Db 486 ATTCTTAAGTCAAGAGAGGAGATTTCCAAAGATATCTACAGAACTTCTGCGAGCATCC 545  
 QY 685 AGAACTGAGTAATAGTCAACTTCTGGCAGACTTCTTCCCTTAATGTTGGGAAACAC 743  
 Db 546 AGAACTGAGTAATAGTCAACTTCTGGCAGACTTCTTCCCTTAATGTTGGGAAACAC 604

RESULT 15  
 BU738240  
 LOCUS BU738240/c 594 bp mRNA linear EST 10-OCT-2002  
 DEFINITION UI-E-DW1-ahc-m-13-0-UI.s1 UI-E-DW1 Homo sapiens cDNA clone  
 UI-E-DW1-ahc-m-13-0-UI 3', mRNA sequence.  
 ACCESSION BU738240  
 VERSION BU738240.1 GI:23675122  
 KEYWORDS EST.  
 SOURCE Homo sapiens (human)  
 ORGANISM Homo sapiens  
 Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 594)  
 AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 97044477  
 PUBMED 8889548  
 COMMENT Contact: Soares, MB  
 Coordinated Laboratory for Computational Genomics  
 University of Iowa  
 375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA  
 Tel: 319 335 8250  
 Fax: 319 335 9565  
 Email: bento-soares@uiowa.edu  
 Tissue Procurement: Dr. Gregg Hageman  
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 Seq primer: M13 FORWARD

FEATURES POLYA=Yes. Location/Qualifiers  
source 1. .594  
/organism="Homo sapiens"  
/mol\_type="mRNA"  
/db\_xref="taxon:9606"  
/clone="UI-E-DWI-ahc-m-13-0-UI"  
/tissue\_type="lens"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/clone\_lib="UI-E-DWI"  
/note="Organ: eye; Vector: pMT3-Pac (Pharmacia) with a modified polylinker; Site 1: EcoR I; Site 2: Not I; UI-E-DWI is a normalized cDNA library containing the following tissue(s): lens. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pMT3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CGATTAGCGA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).  
TAG LIB=UI-E-DWI  
TAG TISSUE=human lens  
TAG\_SEQ=CGATTAGCGA"  
BASE COUNT 184 a 106 c 97 g 207 t  
ORIGIN

Query Match 29.0%; Score 577; DB 13; Length 594;  
Best Local Similarity 100.0%; Pred. No. 7.4e-190;  
Matches 577; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1106 TTAATGGGAACCTGAATCTCTTTAAACACCCCTGGAAATGATAGTACTATCTT 1165  
Db 594 TTAATGGGAACCTGAATCTCTTTAAACACCCCTGGAAATGATAGTACTATCTT 535

QY 1166 CAGGTAAACTAGAACAGCTATTTCAGGACGCGTTGGTCTCCTCATACACTTC 1225  
Db 534 CAGGTAAACTAGAACAGCTATTTCAGGACGCGTTGGTCTCCTCATACACTTC 475

QY 1226 AGAGATGCTATATCTGTGAAACACCTGAACCTCGCTCTCCAGATAAGCAAAAGGA 1285  
Db 474 AGAGATGCTATATCTGTGAAACACCTGAACCTCGCTCTCCAGATAAGCAAAAGGA 415

QY 1286 GCAAAACAGACTTTTGAAGAAATGATGATTAATCCAGATCTGTAGTCAAGTGAT 1345  
Db 414 GCAAAACAGACTTTTGAAGAAATGATGATTAATCCAGATCTGTAGTCAAGTGAT 355

QY 1346 GGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACGCAACAGTA 1405  
Db 354 GGTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACGCAACAGTA 295

QY 1406 CTCACAGCAGCTGACTATGTTTATTCGACATTTGTATACAGGACTGTTTCCAGAG 1465  
Db 294 CTCACAGCAGCTGACTATGTTTATTCGACATTTGTATACAGGACTGTTTCCAGAG 235

QY 1466 CTCATTAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATGTAACACCTGGATTT 1525  
Db 234 CTCATTAAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATGTAACACCTGGATTT 175

QY 1526 GGTATAGATAACCCATGAAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTTTTG 1585  
Db 174 GGTATAGATAACCCATGAAATTTCTGCTGTGCGAGGGTGGTAGAAATTTACTTTTTTG 115

QY 1586 GGTATATTTCTATATATATATATGATACATCGCTGTGCAAAATTTAGTATTTTTCTTTT 1645  
Db 114 GGTATATTTCTATATATATATATGATACATCGCTGTGCAAAATTTAGTATTTTTCTTTT 55

QY 1646 TAATAAGACTAACACAACTTAATGATTAAGTGA 1682

Db 54 TAATAAGACTAACACAACTTAATGATTAAGTGA 18

Search completed: January 31, 2004, 19:32:41  
Job time : 4308 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 31, 2004, 13:56:38 ; Search time 548 Seconds  
(without alignments)  
9812.556 Million cell updates/sec

Title: US-09-744-313A-3  
Perfect score: 1992  
Sequence: 1-gratgaactccaaagt.....tttaataataaaaaaaaa 1992

Scoring table: OLIGO NUC  
Gapop\_60.0 , Gapexb 60.0

Searched: 2552756 seqs 1349719037 residues

Word size: 0  
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :	N_Geneseq_19Jun03.*
1:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25:	/SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1992	100.0	1992	22	SNEXN DNA #1. Hom
2	1838	92.3	2176	22	Human polynucleoti
3	1838	92.3	2176	22	Human polynucleoti
4	1532	76.9	1716	22	Human polynucleoti
5	1296	65.1	3145	24	CDNA encoding a pr
6	918	46.1	1551	22	Human polynucleoti
7	543	27.3	779	20	Human gene express
8	385	19.3	451	24	Breast cancer rela

9	357	17.9	358	21	AAC02755	Human secreted pro
10	78	3.9	580	24	ABN62287	Human cancer relat
11	76	3.8	311	22	AAH69552	Human cervical can
12	76	3.8	313	22	AAH70887	Human cervical can
13	74	3.7	299	22	AAH72445	Human cervical can
14	74	3.7	299	22	AAH73007	Human cervical can
15	60	3.0	60	24	ABN39970	Human spliced tran
16	59	3.0	305	25	ABX50426	Bovine EST associa
17	59	3.0	402	25	ABX50611	Bovine EST associa
18	23	1.2	39380	22	AAK74027	Human immune/haema
19	22	1.1	5807	24	ABL33155	Human immune syste
20	22	1.1	5815	24	ABL70585	Chemically treated
21	22	1.1	6009	24	ABL32829	Human immune syste
22	22	1.1	6070	24	ABQ67129	Human angiogenesis
23	22	1.1	6070	24	ABU70371	Chemically treated
24	22	1.1	6070	24	ABL33678	Human immune syste
25	22	1.1	6070	24	ABL34578	Human metastasis a
26	22	1.1	6277	22	AA546324	Tumour suppressor
27	22	1.1	7049	24	ABL54306	Chemically treated
28	22	1.1	7049	24	ABL32157	Human immune syste
29	22	1.1	7058	24	ABK40052	Human chemically p
30	22	1.1	7238	24	ABL32449	Human immune syste
31	22	1.1	7771	24	ABL33973	Human immune syste
32	22	1.1	8056	25	ABZ10100	Haematopoietic cel
33	22	1.1	8056	25	ABZ10246	Haematopoietic cel
34	22	1.1	8076	24	ABK39955	Human chemically p
35	22	1.1	8576	24	ABL34229	Human immune syste
36	22	1.1	10250	24	ABN80082	Human chemically m
37	22	1.1	10716	24	ABL33419	Human immune syste
38	22	1.1	13511	24	ABL32280	Human immune syste
39	22	1.1	14798	24	ABL33032	Chemically treated
40	22	1.1	19459	24	ABL70527	Chemically treated
41	22	1.1	19459	24	ABK31212	Signal transductio
42	22	1.1	33053	24	ABO67005	Human angiogenesis
43	22	1.1	34548	24	ABL70603	Chemically treated
44	22	1.1	40862	24	ABL34072	Human immune syste
45	22	1.1	73334	24	ABL92318	Chemically treated

# ALIGNMENTS

RESULT 1  
AAC86397  
ID AAC86397 standard; DNA; 1992 BP.

AC AAC86397;  
XX  
DT 21-MAR-2001 (first entry)  
XX  
DE SNEXN DNA #1.

XX SNEXN; human; sorting nexin; inflammation; asthma; allergy; AIDS;  
KW neurological disorder; gastrointestinal; smooth muscle cell;  
KW cancer; gene therapy; ds.

OS Homo sapiens.  
XX  
PN WO200073334-A2.  
XX  
PD 07-DEC-2000.  
XX

XX 26-MAY-2000; 2000WO-US14831.  
XX  
PR 27-MAY-1999; 99US-0136740.  
PR 16-JUN-1999; 99US-0139566.  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
PI Yue H, Tang YT, Azimzai Y;  
XX WPI; 2001-041141/05.  
DR  
XX



PT Novel Sorting Nexin polypeptides and polynucleotides useful for  
 PT diagnosing and treating disorders associated with their expression e.g.  
 PT autoimmune disorders, smooth muscle cell disorders and cell  
 PT proliferative disorders  
 XX Claim 5; Page 82-83; 84pp; English.

XX The present invention relates to human sorting nexin (SNEXN).  
 CC Compositions containing SNEXN or agonists of SNEXN are useful for  
 CC treating a disease or condition associated with decreased expression of  
 CC functional SNEXN and compositions containing antagonists of SNEXN are  
 CC useful for treating a disease or condition associated with  
 CC overexpression of functional SNEXN. These can be immune  
 CC disorders for example inflammation, asthma, allergy, and AIDS,  
 CC neurological disorders, gastrointestinal disorders, smooth muscle cell  
 CC disorders, cancers and others. The SNEXN may also be used for somatic  
 CC or germline gene therapy.

SQ Sequence 1992 BP; 671 A; 308 C; 372 G; 641 T; 0 other;

Query Match 100.0%; Score 1992; DB 22; Length 1992;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1992; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATGAAAATCCAAAAGTTATGTTAGCTATACCTTTAGTTATCACTTTTCAAACTGTT 60  
 DB 1 GTATGAAAATCCAAAAGTTATGTTAGCTATACCTTTAGTTATCACTTTTCAAACTGTT 60

QY 61 TTTCTTTTCTTTTAAATATGTTTTCATATAAAGTACAGCTGTTGATTAATTTGCTATGTA 120  
 DB 61 TTTCTTTTCTTTTAAATATGTTTTCATATAAAGTACAGCTGTTGATTAATTTGCTATGTA 120

QY 121 CTTGATACATTTTGTGTTTAAATATTCAGGAACACACAGAAAAGGGAGATCACTTTGGAAT 180  
 DB 121 CTTGATACATTTTGTGTTTAAATATTCAGGAACACACAGAAAAGGGAGATCACTTTGGAAT 180

QY 181 CAGCAGAAATAGGTAGCAAAATTAAGAGGATTTCAAAGTACACCAATGAGGAGGACTAT 240  
 DB 181 CAGCAGAAATAGGTAGCAAAATTAAGAGGATTTCAAAGTACACCAATGAGGAGGACTAT 240

QY 241 GTTCCCTAAATATGTTGTTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 300  
 DB 241 GTTCCCTAAATATGTTGTTAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 300

QY 301 AATGGAGATGATTTCTCCAGTGGAGGCTGTGAGCAGACACCTTAATCTCCCGAAACCTTGC 360  
 DB 301 AATGGAGATGATTTCTCCAGTGGAGGCTGTGAGCAGACACCTTAATCTCCCGAAACCTTGC 360

QY 361 TGCATGGAAAATTAGCATTCCATATGTAGACTTTTGTAGGATCCCTCTCTGAAAGGAA 420  
 DB 361 TGCATGGAAAATTAGCATTCCATATGTAGACTTTTGTAGGATCCCTCTCTGAAAGGAA 420

QY 421 GGAGAAAAGAAAAGAAATTCCTGTTTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 480  
 DB 421 GGAGAAAAGAAAAGAAATTCCTGTTTGTATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 480

QY 481 AGTTGGACAGAGCTGAACATTTGGTCTGTCTATAGAGATATCTTGAATCTCTACT 540  
 DB 481 AGTTGGACAGAGCTGAACATTTGGTCTGTCTATAGAGATATCTTGAATCTCTACT 540

QY 541 TGAATCAAAATCAACAGAAATTCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 600  
 DB 541 TGAATCAAAATCAACAGAAATTCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 600

QY 601 GATCATTTGGCCCCAAAATATGAATTTCTTAAAGTCAAAGAGGAGGATTCAGAGATA 660  
 DB 601 GATCATTTGGCCCCAAAATATGAATTTCTTAAAGTCAAAGAGGAGGATTCAGAGATA 660

QY 661 TCTACAGAAATCTTCGAGCATCCAGAACTGAGTAATAGTCACTTCTGCGACACTTCT 720  
 DB 661 TCTACAGAAATCTTCGAGCATCCAGAACTGAGTAATAGTCACTTCTGCGACACTTCT 720

QY 721 TTCCCTTAATGGTGGGAAACACAAATTTCTTGATAAGATACCTACAGATGTAATCTTGG 780

DB 721 TTCCCTTAATGGTGGGAAACACAAATTTCTTGATAAGATACCTACAGATGTAATCTTGG 780  
 QY 781 GAAAATTTATAAATCTGTTCTCTGAAAACCTAATGAAAAGAGAAAGGTGAGCAATTTGGAACC 840  
 DB 781 GAAAATTTATAAATCTGTTCTCTGAAAACCTAATGAAAAGAGAAAGGTGAGCAATTTGGAACC 840

QY 841 TTTTATCATGAATTTCAATTAATTTCTGTTGAGTCTCCAAAGCTTAAACCAAGTAGACAGAGA 900  
 DB 841 TTTTATCATGAATTTCAATTAATTTCTGTTGAGTCTCCAAAGCTTAAACCAAGTAGACAGAGA 900

QY 901 ACTGACCATTTCTAGCCCTACTTTTCAGAAAACAAACAAGAGCTTTTCAATCATCTGTTAA 960  
 DB 901 ACTGACCATTTCTAGCCCTACTTTTCAGAAAACAAACAAGAGCTTTTCAATCATCTGTTAA 960

QY 961 AAATAAATGCAAAACCGTCTGAAATATCAGAGAGAGAAAGCAAAATCAGAAATTTATTTGGA 1020  
 DB 961 AAATAAATGCAAAACCGTCTGAAATATCAGAGAGAGAAAGCAAAATCAGAAATTTATTTGGA 1020

QY 1021 GGTGATGACTGTAGAGAGTCTATGANTACCTGATGATGATGATGATGATGATGATGATGATGAT 1080  
 DB 1021 GGTGATGACTGTAGAGAGTCTATGANTACCTGATGATGATGATGATGATGATGATGATGATGAT 1080

QY 1081 GGTTCCTGACTGGCTTCATCATCTCTTAATGGGAACCTCGAATCTCTTTTAAACAACCCCT 1140  
 DB 1081 GGTTCCTGACTGGCTTCATCATCTCTTAATGGGAACCTCGAATCTCTTTTAAACAACCCCT 1140

QY 1141 GGAATGTATGACTGATTAATCTCTTCACTGTAACCTAGAAACAGCTATTTTTCAGGAGCACCG 1200  
 DB 1141 GGAATGTATGACTGATTAATCTCTTCACTGTAACCTAGAAACAGCTATTTTTCAGGAGCACCG 1200

QY 1201 TTTGGTCTCACTATAACACTTTTCAGAGATGCTATATTTCTGTGAAAACACTGAACTGCG 1260  
 DB 1201 TTTGGTCTCACTATAACACTTTTCAGAGATGCTATATTTCTGTGAAAACACTGAACTGCG 1260

QY 1261 CTCTCTCCAGATTAAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGATTAACAT 1320  
 DB 1261 CTCTCTCCAGATTAAGCAAAAGGAGCAAAACAGACTTTTGAAGAAATGATGATTAACAT 1320

QY 1321 TCCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCT 1380  
 DB 1321 TCCAGATCTGTTAGTCAAGTGTATTTGGTGAAGAAACCAAGTATGAAAGCATCAGACTTCT 1380

QY 1381 GTTTGATGGCTTACAGCAACCAAGTCTCAACAGCAGCTGACTATGTTTATTTGGAACAT 1440  
 DB 1381 GTTTGATGGCTTACAGCAACCAAGTCTCAACAGCAGCTGACTATGTTTATTTGGAACAT 1440

QY 1441 TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAGAGAGTTCCTCTGTGAC 1500  
 DB 1441 TGTGATACAGAACTGTTTCCAGAGCTCAATAGGTACAAAGAGAGTTCCTCTGTGAC 1500

QY 1501 ATCTTGGATGTAAACACTTTGGATTTGGTATAGATAACCCATTCGAAATTTCTGCTGTGCG 1560  
 DB 1501 ATCTTGGATGTAAACACTTTGGATTTGGTATAGATAACCCATTCGAAATTTCTGCTGTGCG 1560

QY 1561 AGGGTGGTGAAGAAATTTACTTTTTGGGTATATTTCTTATATATATATATATATATATATATAT 1620  
 DB 1561 AGGGTGGTGAAGAAATTTACTTTTTGGGTATATTTCTTATATATATATATATATATATATATAT 1620

QY 1621 TGAATTTTGTAGTTATTTTGTGTTTAAAGACTAAACAAACTTAAATGATTTAAAGT 1680  
 DB 1621 TGAATTTTGTAGTTATTTTGTGTTTAAAGACTAAACAAACTTAAATGATTTAAAGT 1680

QY 1681 GATTGAGTCTCATAGTCTTTTCAATTTGCTAGCTGTGATCCAAATTTTATTTAGACATAGT 1740  
 DB 1681 GATTGAGTCTCATAGTCTTTTCAATTTGCTAGCTGTGATCCAAATTTTATTTAGACATAGT 1740

QY 1741 CACTTGTATTGCACTTTTAAAGAGAAATTCATATGATGATGATGATGATGATGATGATGATGAT 1800  
 DB 1741 CACTTGTATTGCACTTTTAAAGAGAAATTCATATGATGATGATGATGATGATGATGATGATGAT 1800

QY 1801 GACTGATAAATCTCGTATTTGATAGCTTTGAAAATAAATTTATGCTAGTATGAGAGAAACAG 1860

1801 GACTGATAAAGTTCGATTTGATAGCTTTGAAATAAATATTCCTAGTATGGAGAAACAG 1860  
1861 GAATAAGATCTGATTTCTAGAGTAAATATATTTAGTAGATTTGCTTTCTTTT 1920  
1861 GAATAAGATCTGATTTCTAGAGTAAATATATTTAGTAGATTTGCTTTCTTTT 1920  
1921 ATTTTGTACATAGTTAACTGCTATCTATATAAATAAGCATCTATATGAGTTTATAA 1980  
1921 ATTTTGTACATAGTTAACTGCTATCTATATAAATAAGCATCTATATGAGTTTATAA 1980  
1981 TAAAAAATAA 1992  
1981 TAAAAAATAA 1992  
RESULT 2  
AAI60815/c  
ID AAI60815 standard; cDNA; 2176 BP.  
XX  
AC AAI60815;  
XX  
22-OCT-2001 (first entry)  
DE Human polynucleotide SEQ ID NO 4804.  
XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
KW peripheral nervous system; neuropathy; central nervous system; CNS;  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
KW leukaemia; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200153312-A1.  
XX  
PD 26-JUL-2001.  
XX  
PE 26-DEC-2000; 2000WO-US34263.  
XX  
PR 21-JAN-2000; 2000US-0488725.  
PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042.  
PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450.  
PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX  
PA (HVS8-) HVS8Q INC.  
XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX  
DR WPI: 2001-442253/47.  
DR P-PSDB; AAM1659.  
XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX  
PS Claim 1; SEQ ID NO 4804; 10078pp; English.  
XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC

CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 2176 BP; 684 A; 427 C; 353 G; 712 T; 0 other;  
Query Match 92.3%; Score 1838; DB 22; Length 2176;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1838; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 147 GGACACACAGAAAGGGGAGATCAATTTGGATCAGCAGATAGGTAGCAAAATTAAG 206  
Db 2018 GGACACACAGAAAGGGGAGATCAATTTGGATCAGCAGATAGGTAGCAAAATTAAG 1959  
Qy 207 GAGTATTCAAAGTACCAATCGAGGAGCTATGTTGCTAATTTATGTTAGCTGAGG 266  
Db 1958 GAGTATTCAAAGTACCAATCGAGGAGCTATGTTGCTAATTTATGTTAGCTGAGG 1899  
Qy 267 GTGAAGATGATTTTATTTGAAGAGGTATTTGTTAATGGAAGATGATTTCTCAGTGAGG 326  
Db 1898 GTGAAGATGATTTTATTTGAAGAGGTATTTGTTAATGGAAGATGATTTCTCAGTGAGG 1839  
Qy 327 CTGTGACACACCTAATATCTCCCGAAACCTTCTGCTGATGGAATATAGCATTCATG 386  
Db 1838 CTGTGACACACCTAATATCTCCCGAAACCTTCTGCTGATGGAATATAGCATTCATG 1779  
Qy 387 TAGCATTTTGTAGGATCCCTCTCTGAAAGGAGGAGAGAAAGAAAGAAATTCCTGTGT 446  
Db 1778 TAGCATTTTGTAGGATCCCTCTCTGAAAGGAGGAGAGAAAGAAAGAAATTCCTGTGT 1719  
Qy 447 TTTGTATTGATTTGAAGAAATGATAGAGAGCAGTTGGACACGAGCCTGGAACATTTGT 506  
Db 1718 TTTGTATTGATTTGAAGAAATGATAGAGAGCAGTTGGACACGAGCCTGGAACATTTGT 1659  
Qy 507 CTGTCTATAGAGATATCTTGAATTTCTATGTAATTGAATCAAACTAAACAGATTTTCATG 566  
Db 1658 CTGTCTATAGAGATATCTTGAATTTCTATGTAATTGAATCAAACTAAACAGATTTTCATG 1599  
Qy 567 GTGCATTTCTGATGCGCCAGCTTCTTCTAAGAGGATCATTTGCCCGCCCAAAATTTGAAT 626  
Db 1598 GTGCATTTCTGATGCGCCAGCTTCTTCTAAGAGGATCATTTGCCCGCCCAAAATTTGAAT 1539  
Qy 627 TCTTAAAGTCAAGAGAGGAGAGGTTCCAGAAATATCTACAGAACTTCTCAGCATCCAG 686  
Db 1538 TCTTAAAGTCAAGAGAGGAGAGGTTCCAGAAATATCTACAGAACTTCTCAGCATCCAG 1479  
Qy 687 AACTGAGTAAATAGTCAACTTTCTGCGCAGACTTTCTTCTCCCTAATGTTGGGAAACCAAT 746  
Db 1478 AACTGAGTAAATAGTCAACTTTCTGCGCAGACTTTCTTCTCCCTAATGTTGGGAAACCAAT 1419  
Qy 747 TTTCTGTATAGATATCTACAGATGTAATCTTTGGGAAATATTAATAATCTGTTCTTGAA 806  
Db 1418 TTTCTGTATAGATATCTACAGATGTAATCTTTGGGAAATATTAATAATCTGTTCTTGAA 1359  
Qy 807 AACTAATGAAGAGAGAGGTCAGCATTTGGAACTTTTATCATGAATTTCTTAATTTCTT 866  
Db 1358 AACTAATGAAGAGAGAGGTCAGCATTTGGAACTTTTATCATGAATTTCTTAATTTCTT 1299  
Qy 867 GTGAGTCTCCAAAGCCTAAACCAAGTAGACAGCACTGACCATTTCTCAGCCCTACTTCAG 926  
Db 1298 GTGAGTCTCCAAAGCCTAAACCAAGTAGACAGCACTGACCATTTCTCAGCCCTACTTCAG 1239  
Qy 927 AAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGGAATA 986  
Db 1238 AAAACAACAAGAGCTTTTCAATGATCTGTTTAAATAATGCAAAACCGTCTGGAATA 1179  
Qy 987 CAGAGAGAAACCAAAATCAGAAATTTATTTATGAGGATGATCTGTAGAGAGTCTATG 1046  
Db 1178 CAGAGAGAAACCAAAATCAGAAATTTATTTATGAGGATGATCTGTAGAGAGTCTATG 1119





PR 19-OCT-2000; 2000US-0693036.  
PR 29-NOV-2000; 2000US-0727344.  
XX (HYSE-) HYSEQ INC.  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
XX WPI; 2001-442253/47.  
DR P-PSDB; AAM39873.  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
XX Claim 1; SEQ ID NO 1232; 10078pp; English.  
XX The invention relates to human nucleic acids (AA157798-AA161369) and  
CC the encoded polypeptides (AAM38642-AAW42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
SQ Sequence 1716 BP; 572 A; 297 C; 347 G; 500 T; 0 other;  
  
Query Match 76.98; Score 1532; DB 22; Length 1716;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1532; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 147 GGAACACACAGAAAGGGAGATCAATTTGGATTCAGACAGATAGGTAGCMAAATTAAG 206  
DB 177 GGAACACACAGAAAGGGAGATCAATTTGGATTCAGACAGATAGGTAGCMAAATTAAG 236  
QY 207 GAGTATTTCAAAGTACACAAATGAGGAGCTATGTTGCCATTAATGATGTTAGTGAAG 266  
DB 237 GAGTATTTCAAAGTACACAAATGAGGAGCTATGTTGCCATTAATGATGTTAGTGAAG 296  
QY 267 GTGAAGATGATTTATTTGAAGAGATGTTGTTATGTAAGATGATTTCCAGTGGAGG 326  
DB 297 GTGAAGATGATTTATTTGAAGAGATGTTGTTATGTAAGATGATTTCCAGTGGAGG 356  
QY 327 CTGTGACACACCTAATTAATCTCCCGAACCTTGTGCTGATGAAATTTAGCATTTCCATATG 386  
DB 357 CTGTGACACACCTAATTAATCTCCCGAACCTTGTGCTGATGAAATTTAGCATTTCCATATG 416  
QY 387 TAGACTTTTGTGAGATCCCTCTCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 446  
DB 417 TAGACTTTTGTGAGATCCCTCTCTGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 476  
QY 447 TTTGATTTGATGTTGAAG 506  
DB 477 TTTGATTTGATGTTGAAG 536  
QY 507 CTGTCTATAGAGATATCTTGAATTTATGATGATGATGATGATGATGATGATGATGATGATG 566  
DB 537 CTGTCTATAGAGATATCTTGAATTTATGATGATGATGATGATGATGATGATGATGATGATG 596  
QY 567 GTGCATTTCTGATGCCAGCTTCTCTTAAGAGAGATCATTTGGCCGCCCAAAAATTATGAAT 626  
DB 597 GTGCATTTCTGATGCCAGCTTCTCTTAAGAGAGATCATTTGGCCGCCCAAAAATTATGAAT 656  
QY 627 TCTTAAAGTCAAAGAGGAGAGATTTCCAGAAATATCTACAGAAATCTTCTGCAGCATCCAG 686

DB 657 TCTTAAAGTCAAAGAGGAGAGATTTCCAGAAATATCTACAGAAATCTTCTGCAGCATCCAG 716  
QY 687 AACTGAGTAAATAGTCAACTTCTGGCAGACTTTTCCCTAATGCTGGGAAACCAAT 746  
DB 717 AACTGAGTAAATAGTCAACTTCTGGCAGACTTTTCCCTAATGCTGGGAAACCAAT 776  
QY 747 TCTTTGATTAAGATACTACAGATGTAATCTTGGGAAATTAATAAATCTGTTCTCTGGAA 806  
DB 777 TCTTTGATTAAGATACTACAGATGTAATCTTGGGAAATTAATAAATCTGTTCTCTGGAA 836  
QY 807 AACTTAATGAAGAGAAAGGTGAGCAATTTGGAACCTTTTATCATGAATTTCAATTAATCTT 866  
DB 837 AACTTAATGAAGAGAAAGGTGAGCAATTTGGAACCTTTTATCATGAATTTCAATTAATCTT 896  
QY 867 GTGAGTCTCAAAGCCTAAACCAAGTAGACAGACCTGACCATTTCTCAGCCCTACTCTAG 926  
DB 897 GTGAGTCTCAAAGCCTAAACCAAGTAGACAGACCTGACCATTTCTCAGCCCTACTCTAG 956  
QY 927 AAAACCAACAGAGACTTTTCAATGATCTGTTTAAAAAATGCAAAACCGTCTGAAATA 986  
DB 957 AAAACCAACAGAGACTTTTCAATGATCTGTTTAAAAAATGCAAAACCGTCTGAAATA 1016  
QY 987 CAGAGAGAAAGCAAAATCAGAAATTTATTTATGAGGATGATGATGATGAGAGAGTCTATG 1046  
DB 1017 CAGAGAGAAAGCAAAATCAGAAATTTATTTATGAGGATGATGATGATGAGAGAGTCTATG 1076  
QY 1047 ATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1106  
DB 1077 ATTAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1136  
QY 1107 TAATGGGAATCGAATCTCTTTAAAAACACCCCTGGAAATGATGATGATGATGATGATGATG 1166  
DB 1137 TAATGGGAATCGAATCTCTTTAAAAACACCCCTGGAAATGATGATGATGATGATGATGATG 1196  
QY 1167 AGTGTAACACTAGAACAGCTATTTTCCAGGAGCAGCTTTTGGTCTCCTCCTCACTAACACTCTCA 1226  
DB 1197 AGTGTAACACTAGAACAGCTATTTTCCAGGAGCAGCTTTTGGTCTCCTCCTCACTAACACTCTCA 1256  
QY 1227 GAGATGCTATATTTCTGTGAAAAACACTGAACTGCTGCTCTCTCCAGAGATGAGCAAAAGGAG 1286  
DB 1257 GAGATGCTATATTTCTGTGAAAAACACTGAACTGCTGCTCTCTCCAGAGATGAGCAAAAGGAG 1316  
QY 1287 CAAAACAGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1346  
DB 1317 CAAAACAGACTTTTGAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1376  
QY 1347 GTGAGAAACCAAGTATGAAAGCATCAGACTTCTGTTTGGCTTATGCTTATGCTTATGCTTATG 1406  
DB 1377 GTGAGAAACCAAGTATGAAAGCATCAGACTTCTGTTTGGCTTATGCTTATGCTTATGCTTATG 1436  
QY 1407 TCAAACAGCAGCTGACTTATGTTTATTTGGAATTTGGAATTTGGAATTTGGAATTTGGAATTTG 1466  
DB 1437 TCAAACAGCAGCTGACTTATGTTTATTTGGAATTTGGAATTTGGAATTTGGAATTTGGAATTTG 1496  
QY 1467 TCAAATAGGTTCAAAAGAGTTACCTCTGTGCAATCTTGGATGTAACACTTGGATTTG 1526  
DB 1497 TCAAATAGGTTCAAAAGAGTTACCTCTGTGCAATCTTGGATGTAACACTTGGATTTG 1556  
QY 1527 GTATAGATTAACCAATGAAATTTCTGCTGCGAGGGGTGAGAAATTTACTTTTTTGG 1586  
DB 1557 GTATAGATTAACCAATGAAATTTCTGCTGCGAGGGGTGAGAAATTTACTTTTTTGG 1616  
QY 1587 GTATATTTCTTATATATATTTATGATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1646  
DB 1617 GTATATTTCTTATATATATTTATGATCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1676  
QY 1647 AATAAGACTAACCAAACTTAATTAATAA 1678  
DB 1677 AATAAGACTAACCAAACTTAATTAATAA 1708

ABQ79518 standard; cDNA; 3145 BP.

XX AC ABQ79518;

XX DT 25-NOV-2002 (first entry)

XX XX cDNA encoding a protein similar to human sorting nexin.

XX DE Atherosclerosis; antiarteriosclerotic; marker; cardiovascular;

XX KW SSH 6; gene; human; nexin; ss.

XX OS Homo sapiens.

XX XX Key Location/Qualifiers

XX FH 128..2968

XX CDS /\*tag= a

XX FT

XX PN W0200262839-A2.

XX XX 15-AUG-2002.

XX XX 05-FEB-2002; 2002WO-EP01327.

XX PF 07-FEB-2001; 2001BP-0200439.

XX PR (UYMA-) UNIV MAASTRICHT.

XX PA Daemen MJAP, Cleutjens CBJM, Zaman GJR;

XX PI WPI; 2002-643400/69.

XX DR P-PSDB; ABB81193.

XX XX

XX PT Use of a polynucleotide differentially expressed in ruptured and stable

XX PT atherosclerotic plaques as a marker for atherosclerosis, useful in

XX PT treating, diagnosing or preventing atherosclerosis -

XX XX

XX XX Example 6; Page 35-37; 44pp; English.

XX CC The invention relates to the use of a polynucleotide differentially

XX CC expressed in ruptured and stable atherosclerotic plaques as a marker

XX CC for atherosclerosis, where the polynucleotides can be selected from the

XX CC sequences shown in ABQ79517-19. The polynucleotides are useful as a

XX CC marker of atherosclerosis, which may be used: (i) in the diagnosis,

XX CC prevention and treatment of atherosclerosis; (ii) as serum/plasma markers

XX CC to screen patients at risk for plaque instability to evaluate the effects

XX CC of other treatments; (iii) in the preparation of vector molecules for the

XX CC expression of the encoded protein in host cells; and (iv) in the

XX CC identification of functional targets or analogues of the gene. The

XX CC polynucleotides, the encoded proteins or antibodies against the proteins

XX CC may be used to target other therapeutics to an unstable plaque.

XX CC Modulation of the expression of the polynucleotide can increase plaque

XX CC stability and therefore inhibit the progression of atherosclerotic

XX CC cardiovascular disease. Modulators may be used to prepare pharmaceuticals

XX CC for atherosclerotic disorders. The present sequence represents a cDNA

XX CC encoding a protein similar to human sorting nexin.

XX XX

XX SQ Sequence 3145 BP; 985 A; 570 C; 659 G; 931 T; 0 other;

Query Match 65.1%; Score 1296; DB 24; Length 3145;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1446; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 147 GGAACACACAGAAAGGGGAGATCATTTGGATCAGCAGATAGGTAGCAAAATTAAG 206

DB 1602 GGAACACACAGAAAGGGGAGATCATTTGGATCAGCAGATAGGTAGCAAAATTAAG 1661

QY 207 GAGTATTCAAAGTACCACAAATGGAGGAGCTATGTTGCTTAATTATGGTGTAGCTGAAG 266

DB 1662 GAGTATTCAAAGTACCACAAATGGAGGAGCTATGTTGCTTAATTATGGTGTAGCTGAAG 1721

QY 267 GTGAAGATGATTTTATTGAAGAAGAGTATTGTTGTAATGGAAGATGATTTCCAGTGGAGG 326



QY 1407 TCAACAGCAGCTGACCTTATGTTTATGACATCTGTGATACAGCACTGTTCCAGAC 1466  
DB 2862 TCAACAGCAGCTGACCTTATGTTTATGACATCTGTGATACAGCACTGTTCCAGAC 2921  
QY 1467 TCAATAGGTACAAAAGGAGTCTCTGTGACATCTTGGATGTAAACACTTGGATTG 1526  
DB 2922 TCAATAGGTACAAAAGGAGTCTCTGTGACATCTTGGATGTAAACACTTGGATTG 2981  
QY 1527 GTATAGAAATACCCATTGAAATTTCTGTGCGAGGGTGTAGAAATTTACTTTTGG 1586  
DB 2982 GTATAGAAATACCCATTGAAATTTCTGTGCGAGGGTGTAGAAATTTACTTTTGG 3041  
QY 1587 GTATATTCT 1595  
DB 3042 GTATATTCT 3050

## RESULT 6

AAI59030  
ID AAI59030 standard; cDNA; 1551 BP.

AC AAI59030;

XX 22-OCT-2001 (first entry)

DT Human polynucleotide SEQ ID NO 1233.

XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
XX peripheral nervous system; neuropathy; central nervous system; CNS;  
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
XX anyotropic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
XX leukaemia; ss.

XX Homo sapiens.

XX WO200153312-A1.

XX 26-JUL-2001.

XX 26-DEC-2000; 2000WO-US34263.

XX 21-JAN-2000; 2000US-0488725.

XX 25-APR-2000; 2000US-0552317.

XX 09-JUL-2000; 2000US-0598042.

XX 19-JUL-2000; 2000US-0620312.

XX 03-AUG-2000; 2000US-0653450.

XX 14-SEP-2000; 2000US-0662191.

XX 19-OCT-2000; 2000US-0693036.

XX 29-NOV-2000; 2000US-0727344.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX WPI; 2001-442253/47.

XX P-PSDB; AAM39874.

XX Novel nucleic acids and polypeptides, useful for treating disorders

XX such as central nervous system injuries -

XX Claim 1; SEQ ID NO 1233; 10078pp; English.

XX The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,

XX immunosuppressant and cytostatic activity. The polynucleotides are useful

XX in gene therapy. A composition containing a polypeptide or polynucleotide

XX of the invention may be used to treat diseases of the peripheral nervous

XX system, such as peripheral nervous injuries, peripheral neuropathy and

XX localised neuropathies and central nervous system diseases, such as

CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemia and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.

SQ Sequence 1551 BP; 527 A; 258 C; 317 G; 449 T; 0 other;

Query Match 46.1%; Score 918; DB 22; Length 1551;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 GGAACACACAGAAAAGGGAGAAATCAATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 206

DB 177 GGAACACACAGAAAAGGGAGAAATCAATTTGGAATCAGCAGATAGGTAGCAAAATTAAG 236

QY 207 GAGTATTCAAAAGTACCAATCGAGGAGCTATGTTGCTTAAATTTATGTTAGCTGAAG 266

DB 237 GAGTATTCAAAAGTACCAATCGAGGAGCTATGTTGCTTAAATTTATGTTAGCTGAAG 296

QY 267 GTGAGATGATTTTATTTGAAGAAGGTATTTGTTGTAATGGAAGATGATTCCTCAGTGGAG 326

DB 297 GTGAAGATGATTTTATTTGAAGAAGGTATTTGTTGTAATGGAAGATGATTCCTCAGTGGAG 356

QY 327 CTGTGACACACCTTAATACCTCCCGAAACCTCTGTCATCGAAATTTAGCATTTCCATATG 386

DB 357 CTGTGACACACCTTAATACCTCCCGAAACCTCTGTCATCGAAATTTAGCATTTCCATATG 416

QY 387 TAGACTTTTGTGAGATCCCTCTCTGAAAGGAGGAGAAAAGAAAGAAATTCCTGTGT 446

DB 417 TAGACTTTTGTGAGATCCCTCTCTGAAAGGAGGAGAAAAGAAAGAAATTCCTGTGT 476

QY 447 TTGTGATGATTTGAAGAAGATGATAGAGACGCTTGACACGCGCTGGAACATTTGT 506

DB 477 TTGTGATGATTTGAAGAAGATGATAGAGACGCTTGACACGCGCTGGAACATTTGT 536

QY 507 CTGTCTATAGAGATATCTTGAATCTTATGTTGTAATCAAAATCAACAGAAATTTATG 566

DB 537 CTGTCTATAGAGATATCTTGAATCTTATGTTGTAATCAAAATCAACAGAAATTTATG 596

QY 567 GTGCATTTCTGTATGCCAGCTTCTTCTAAGAGGATCATTTGCCCCCAAAAATTTAAT 626

DB 597 GTGCATTTCTGTATGCCAGCTTCTTCTAAGAGGATCATTTGCCCCCAAAAATTTAAT 656

QY 627 TCTTAAAGTCAAGAGGAGGAGAGGTCCAGAAATATCTACAGAAATCTTCGAGCATCCAG 686

DB 657 TCTTAAAGTCAAGAGGAGGAGAGGTCCAGAAATATCTACAGAAATCTTCGAGCATCCAG 716

QY 687 AACTGATATAGTCAACTTCTGCGAGACTTTCTTCCCTTAATGTTGGGAAACACAAT 746

DB 717 AACTGATATAGTCAACTTCTGCGAGACTTTCTTCCCTTAATGTTGGGAAACACAAT 776

QY 747 TCTTGTATAGATATACTACAGATGTAATCTTGGGAAATTTAATAATCTGTTCTGGAA 806

DB 777 TCTTGTATAGATATACTACAGATGTAATCTTGGGAAATTTAATAATCTGTTCTGGAA 836

QY 807 AACTAATGAAGAGAGAGGTGAGCATTTGGAACCTTTTATCATGAATTTCAATTAATCTT 866

DB 837 AACTAATGAAGAGAGAGGTGAGCATTTGGAACCTTTTATCATGAATTTCAATTAATCTT 896

QY 867 GTGATCTCCAAAGCCTTAAACCAAGTAGACAGACCTGACCATCTCAGCCCTACTTCAG 926

DB 897 GTGATCTCCAAAGCCTTAAACCAAGTAGACAGACCTGACCATCTCAGCCCTACTTCAG 956

QY 927 AAAACACACAGAACTTTTCAATGATCTGTTTAAAAAATAATGAAACCGTCTGAAATA 986

DB 957 AAAACACACAGAACTTTTCAATGATCTGTTTAAAAAATAATGAAACCGTCTGAAATA 1016

QY 987 CAGAGAGAAACAAAATCAGAAATTTATTTATGAGGTGATGACTGTAGAGAGTCTATG 1046

XX	Sequence 779 BP; 248 A; 143 C; 153 G; 232 T; 3 other;
SQ	Query Match 27.3%; Score 543; DB 20; Length 779; Best Local Similarity 100.0%; Pred. No. 8.8e-237; Matches 543; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	917 CCTACTTCAGAAAACCAAGAGCTTTTCAATGATCTGTTTAAAAAATATGCAAAACCGT 976
DB	95 CCTACTTCAGAAAACCAAGAGCTTTTCAATGATCTGTTTAAAAAATATGCAAAACCGT 154
QY	977 GCTGAAAATACAGAGAGAAAACCAAGATTTATTTATGAGGTGATGCTGTAGAA 1036
DB	155 GCTGAAAATACAGAGAGAAAACCAAGATTTATTTATGAGGTGATGCTGTAGAA 214
QY	1037 GGAGTCTATGATTTACCTGATGTATGAGGCGGTAGTTTCCAGGTTCTCGACTGCTT 1096
DB	215 GGAGTCTATGATTTACCTGATGTATGAGGCGGTAGTTTCCAGGTTCTCGACTGCTT 274
QY	1097 CATCATCTCTTAATGGGAACCTCGAATCTCTTTAAAAAACACCCCTGGAAAATGTATCTGAT 1156
DB	275 CATCATCTCTTAATGGGAACCTCGAATCTCTTTAAAAAACACCCCTGGAAAATGTATCTGAT 334
QY	1157 TACTATCTTCAGTGTAAACTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTCACTCATA 1216
DB	335 TACTATCTTCAGTGTAAACTAGAACAGCTATTTTCAGGAGCACCGTTTGGTCTCACTCATA 394
QY	1217 ACACCTTCTCAGAGATGCTATTTCTGTGAAAACCACTGAACCTCGCTCTCTCCAGATAAG 1276
DB	395 ACACCTTCTCAGAGATGCTATTTCTGTGAAAACCACTGAACCTCGCTCTCTCCAGATAAG 454
QY	1277 CAAAAAGGAGCAAAACAGACTTTTGAAGAATGATGATTTACATTCAGATCTGTAGTC 1336
DB	455 CAAAAAGGAGCAAAACAGACTTTTGAAGAATGATGATTTACATTCAGATCTGTAGTC 514
QY	1337 AAGTGATTTGTTGTAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTTGTGCTTACAG 1396
DB	515 AAGTGATTTGTTGTAAGAAACCAAGTATGAAAGCATCAGACTTCTGTTTGTGCTTACAG 574
QY	1397 CAACCACTACTCAACAGCAGCTGACTTATGTTTATTTGAGCATTTGTGATACAGAACTG 1456
DB	575 CAACCACTACTCAACAGCAGCTGACTTATGTTTATTTGAGCATTTGTGATACAGAACTG 634
QY	1457 TTT 1459
DB	635 TTT 637
RESULT 8	
ABL63556/c	
ID	ABL63556 standard; DNA; 451 BP.
XX	ABL63556;
AC	
DT	15-MAY-2002 (first entry)
DE	Breast cancer related gene sequence SEQ ID NO:1893.
XX	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW	stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW	cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW	gene; ds.
XX	Homo sapiens.
OS	
XX	WO200194629-A2.
PN	
XX	13-DEC-2001.
PD	
XX	30-MAY-2001; 2001WO-US10838.
PF	
XX	05-JUN-2000; 2000US-209473P.
PR	05-JUN-2000; 2000US-209531P.

Db	1017	CAGAGAGAAAGCAATATCAGATTTATTTATGAGGTGATGACTGTAGAGGACTATG	1076
Qy	1047	ATTACCTGATGTATGTAG	1064
Db	1077	ATTACCTGATGTATGTAG	1094
RESULT 7			
XX	AAZ17379		
ID	AAZ17379	standard; cDNA; 779 BP.	
AC	AAZ17379;		
XX			
DT	12-OCT-1999	(first entry)	
DE	Human gene expression product cDNA sequence SEQ ID NO:4852.		
XX	Human; gene; gene expression product; diagnosis; therapy; probe;		
KW	detection; mapping; tissue typing; profiling; forensic; cancer;		
KW	genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.		
XX	Homo sapiens.		
XX	WO9938972-A2.		
XX	05-AUG-1999.	99WO-US01619.	
XX	28-JAN-1999;		
XX	03-APR-1998;	98US-0080666.	
PR	28-JAN-1998;	98US-0072910.	
PR	24-FEB-1998;	98US-0075954.	
PR	31-MAR-1998;	98US-0080114.	
PR	03-APR-1998;	98US-0080515.	
XX	(CHIR ) CHIRON CORP.		
PA	(HYSE-) HYSEQ INC.		
XX	Crkvenjakov R, Dickson M, Drmanac R, Drmanac S;		
PI	Escobedo J, Garcia PD, Garcia V, Giese K, Innis MA;		
PI	Jones WL, Kassam A, Kennedy GC, Kita D, Labat I;		
PI	Lamson G, Leshkowitz D, Pot D, Randazzo F, Reinhard C;		
PI	Stache-Crain B, Sudduth-Klinger J, Williams LT;		
XX	WPI; 1999-494092/41.		
XX	Novel human genes and their expression products which are		
PT	differentially expressed in different cell types		
XX	Claim 1; Page 2307; 2479pp; English.		
XX	The present invention describes a library of human polynucleotides		
CC	comprising the sequences given in AAZ12532 to AAZ17779. Also described is		
CC	a method of detecting differentially expressed genes correlated with the		
CC	cancerous state of a mammalian cell, comprising detecting at least one		
CC	differentially expressed gene product in a test sample from a cell		
CC	suspected of being cancerous, where the gene product is encoded by one		
CC	of the 5248 polynucleotide sequences given in AAZ12532 to AAZ17779. The		
CC	polynucleotides can be used as a source of primers and probes, which can		
CC	be used for a variety of purpose, e.g. detection of expression levels,		
CC	mapping, tissue typing or profiling, forensics, genetic analysis and		
CC	detection of polymorphisms. Polypeptides encoded by the polynucleotides		
CC	can be used for raising antibodies for experimental, diagnostic and		
CC	therapeutic purposes. The polynucleotides may also be used to construct		
CC	arrays for diagnostics (which may be used to determine function of an		
CC	encoded protein); and to detect differences in expression levels between		
CC	two cells (e.g. to identify abnormal or diseased tissue in a human, to		
CC	identify a genetic predisposition or susceptibility to a disease such as		
CC	cancer). The polynucleotides of the invention are especially used in the		
CC	diagnosis, prognosis and management of colorectal cancer, breast cancer,		
CC	and lung cancer. The polynucleotides can also be used to screen for		
CC	peptide analogues and antagonists.		

18-SEP-2000; 2000US-233133P.  
 PR 18-SEP-2000; 2000US-233167P.  
 PR 20-SEP-2000; 2000US-234009P.  
 PR 20-SEP-2000; 2000US-234034P.  
 PR 20-SEP-2000; 2000US-234052P.  
 PR 22-SEP-2000; 2000US-234509P.  
 PR 22-SEP-2000; 2000US-234567P.  
 PR 25-SEP-2000; 2000US-234923P.  
 PR 25-SEP-2000; 2000US-234924P.  
 PR 25-SEP-2000; 2000US-235077P.  
 PR 25-SEP-2000; 2000US-235082P.  
 PR 25-SEP-2000; 2000US-235113P.  
 PR 26-SEP-2000; 2000US-235280P.  
 PR 26-SEP-2000; 2000US-235637P.  
 PR 26-SEP-2000; 2000US-235638P.  
 PR 27-SEP-2000; 2000US-235711P.  
 PR 27-SEP-2000; 2000US-235720P.  
 PR 27-SEP-2000; 2000US-235840P.  
 PR 27-SEP-2000; 2000US-235863P.  
 PR 28-SEP-2000; 2000US-236028P.  
 PR 28-SEP-2000; 2000US-236032P.  
 PR 28-SEP-2000; 2000US-236033P.  
 PR 28-SEP-2000; 2000US-236034P.  
 PR 28-SEP-2000; 2000US-236109P.  
 PR 28-SEP-2000; 2000US-236111P.  
 PR 29-SEP-2000; 2000US-236842P.  
 PR 29-SEP-2000; 2000US-236891P.  
 PR 02-OCT-2000; 2000US-237172P.  
 PR 02-OCT-2000; 2000US-237173P.  
 PR 02-OCT-2000; 2000US-237278P.  
 PR 02-OCT-2000; 2000US-237294P.  
 PR 02-OCT-2000; 2000US-237295P.  
 PR 02-OCT-2000; 2000US-237316P.  
 PR 02-OCT-2000; 2000US-237318P.  
 PR 03-OCT-2000; 2000US-237598P.  
 PR 03-OCT-2000; 2000US-237604P.  
 PR 03-OCT-2000; 2000US-237605P.  
 PR 03-OCT-2000; 2000US-237608P.  
 PR 01-NOV-2000; 2000US-244867P.  
 PR 01-NOV-2000; 2000US-245084P.  
 PA (AVAL-) AVALON PHARM.  
 XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
 PI Soppet DR, Weaver Z;  
 XX WPI; 2002-188264/24.  
 XX Screening for anti-neoplastic agent involves exposing cells to a  
 PT chemical agent to be tested for anti-neoplastic activity, and  
 PT determining a change in expression of a gene of a signature gene set -  
 XX Claim 1; SEQ ID 1893; 44pp; English.  
 XX The present invention describes a method (M1) for screening for an  
 CC anti-neoplastic agent. The method involves exposing cells to a chemical  
 CC agent to be tested for anti-neoplastic activity, determining a change in  
 CC expression of at least one gene (I) of a signature gene set, where (I)  
 CC comprises a sequence (S) selected from 8447 sequences given in ABL61664  
 CC to ABL70110, or is at least 95% identical to (S), where a change in  
 CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
 CC activity and can be used in gene therapy. M1 can be used for screening  
 CC an anti-neoplastic agent, and can be used for producing a product which  
 CC is the data collected with respect to the anti-neoplastic agent as a  
 CC result of M1, and the data is sufficient to convey the chemical  
 CC structure and/or properties of the agent. M1 can be used in the  
 CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
 CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
 CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
 CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
 CC carcinoma, papillary carcinoma and Wilms' tumour.  
 XX Sequence 451 BP; 179 A; 73 C; 52 G; 146 T; 1 other;  
 SQ

Query Match 19.3%; Score 385; DB 24; Length 451;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-165;  
 Matches 385; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1597 ATATATATATGACATCGCTGCTGAAATTTAGTATATTTTGTATTAAGAGACT 1656  
 DB 385 ATATATATATGACATCGCTGCTGAAATTTAGTATATTTTGTATTAAGAGACT 326  
 QY 1657 AACACAACTTAATGATTAAGATGAGTCTCATAGTCTTTTCATTGCTAGCTGTGA 1716  
 DB 325 AACACAACTTAATGATTAAGATGAGTCTCATAGTCTTTTCATTGCTAGCTGTGA 266  
 QY 1717 TCCAAATTTTATAGAACATAAGTCACTGCTGTTATGCGCATTTTAAAGAGAAAATTCAT 1776  
 DB 265 TCCAAATTTTATAGAACATAAGTCACTGCTGTTATGCGCATTTTAAAGAGAAAATTCAT 206  
 QY 1777 AATGATGTTATGCGCAACAGATGAGTCTGATAAAGTCTGATTTGTTATGTTAGTAAATA 1836  
 DB 205 AATGATGTTATGCGCAACAGATGAGTCTGATAAAGTCTGATTTGTTATGTTAGTAAATA 146  
 QY 1837 ATTATGCTAGTATGAGGAGAACAGGATAAGATCTGATTTTCTTAGAGTTAATATATTTT 1896  
 DB 145 ATTATGCTAGTATGAGGAGAACAGGATAAGATCTGATTTTCTTAGAGTTAATATATTTT 86  
 QY 1897 AGTAGATTGTTTTCCTTTTATTTTATTTTGTACATAGTAACTGTGTATCTATAATAA 1956  
 DB 85 AGTAGATTGTTTTCCTTTTATTTTATTTTGTACATAGTAACTGTGTATCTATAATAA 26  
 QY 1957 GCATCCTATATGAGTTTATTAAT 1981  
 DB 25 GCATCCTATATGAGTTTATTAAT 1

RESULT 9  
 AAC02755  
 ID AAC02755 standard; cDNA; 358 BP.  
 XX AAC02755;  
 XX 06-OCT-2000 (first entry)  
 XX Human secreted protein 5' EST, SEQ ID NO: 2753.  
 XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping; ss.  
 XX Homo sapiens.  
 XX EP1033401-A2.  
 XX 06-SEP-2000.  
 XX 21-FEB-2000; 2000EP-0200610.  
 XX 26-FEB-1999; 99US-0122487.  
 XX (GEST) GENSET.  
 XX Dumas Milne Edwards J, Duclert A, Giordano J;  
 XX WPI; 2000-500381/45.  
 XX P-PSDB; AAG02749.  
 XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures -  
 PS Claim 1; SEQ ID 2753; 71pp + CD-ROM; English.  
 CC The present sequence is one of a large number of 5' ESTs derived from  
 CC mRNAs encoding secreted proteins. An ORF has been identified within the  
 CC sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs

CC derived from 30 different tissues. EST sequences usually correspond  
CC mainly to the 3' untranslated region (UTR) of the mRNA because they are  
CC often obtained from oligo-dT primed cDNA libraries. Such ESTs are not  
CC well suited for isolating cDNA sequences derived from the 5' ends of  
CC mRNAs and even in those cases where longer cDNA sequences have been  
CC obtained, the full 5' UTR is rarely included. 5' ESTs are derived from  
CC cDNAs with intact 5' ends and can therefore be used to obtain full length  
CC cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,  
CC gene therapy and chromosome mapping procedures. They are used to obtain  
CC upstream regulatory sequences and to design expression and secretion  
CC vectors.

XX Sequence 358 BP; 125 A; 63 C; 69 G; 101 T; 0 other;

Query Match 17.9%; Score 357; DB 21; Length 358;

Best Local Similarity 100.0%; Pred. No. 3.8e-152;

Matches 357; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 760 ACTACAGATGTAATCTTTGGGAAATATATAAATCTGTCTCGAAACTAATGAAGA 819  
DB 1 ACTACAGATGTAATCTTTGGGAAATATATAAATCTGTCTCGAAACTAATGAAGA 60

QY 820 GAAAGTCAGACCTTTGGAACTTTTATCATGAATTTCAATTAATCTTGTGAGTCTCCAAA 879  
DB 61 GAAAGTCAGACCTTTGGAACTTTTATCATGAATTTCAATTAATCTTGTGAGTCTCCAAA 120

QY 880 GCGTAAACCAAGTAGACAGACTGACCAATTCAGCCCTACTTCAGAAAACACAGAA 939  
DB 121 GCGTAAACCAAGTAGACAGACTGACCAATTCAGCCCTACTTCAGAAAACACAGAA 180

QY 940 GCTTTTCATGATCTGTTTAAATAATGCAACCGTCTGAAATACAGAGAGAAGCA 999  
DB 181 GCTTTTCATGATCTGTTTAAATAATGCAACCGTCTGAAATACAGAGAGAAGCA 240

QY 1000 AAATCAGATTTATTTATGAGTGATGATCTAGAGAGTCTATGATTACCTGATGTA 1059  
DB 241 AAATCAGATTTATTTATGAGTGATGATCTAGAGAGTCTATGATTACCTGATGTA 300

QY 1060 TGTAGACGGGTAGTTTTCCAGGTTCTGACTGGCTTCATCATCTCTTAATGGGAAC 1116

DB 301 TGTAGACGGGTAGTTTTCCAGGTTCTGACTGGCTTCATCATCTCTTAATGGGAAC 357

RESULT 10

ABN62287/c

ID ABN62287 standard; cDNA; 580 BP.

XX AC ABN62287;

XX DT 28-JUN-2002 (first entry)

XX DE Human cancer related polynucleotide SEQ ID NO 2254.

XX KW Human; cytostatic; gene expression; gene mapping; tissue profiling;

XX KW gene therapy; cancer; tumour; gene; ss.

XX OS Homo sapiens.

XX PN WO200214500-A2.

XX PD 21-FEB-2002.

XX PF 16-AUG-2001; 2001WO-US25840.

XX PR 16-AUG-2000; 2000US-226326P.

XX PA (CHIR ) CHIRON CORP.

XX PA (HYSE-) HYSEQ INC.

XX PI Escobedo J, Garcia PD, Sudduth-Klinger J, Reinhard C, Randazzo F;

XX PI Lamson G, Scott EW, Zhang G, Kassam A, Pot D, Labat I;

XX DR WPI; 2002-241905/29.

XX New nucleic acid for producing a polypeptide, detecting differentially  
XX expressed genes correlated with a cancerous state of a mammalian cell,  
XX PT and inhibiting tumor growth -

XX PS Claim 1; SEQ ID NO 2254; 883pp + Sequence Listing; English.

XX The invention relates to an isolated polynucleotide (ABN27253-ABN33262)  
XX with cytostatic activity. The polynucleotide is used to produce a  
XX polypeptide, to detect differentially expressed genes correlated with a  
XX cancerous state of a mammalian cell and to inhibit tumour growth. The  
XX polynucleotide is used as a probe in mapping and tissue profiling. The  
XX encoded polypeptide and antibodies to the polypeptide can also be used  
XX for therapeutic and diagnostic purposes. The polynucleotide is useful for  
XX gene therapy.

XX Note: The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 580 BP; 242 A; 78 C; 94 G; 166 T; 0 other;

Query Match 3.9%; Score 78; DB 24; Length 580;

Best Local Similarity 100.0%; Pred. No. 3.3e-25;

Matches 78; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTATGAACCTCCAAAGTTATGTTAGCTATACCTTTAGTTTATCATCTTTCAAACTGTT 60

DB 86 GTATGAACCTCCAAAGTTATGTTAGCTATACCTTTAGTTTATCATCTTTCAAACTGTT 27

QY 61 TTTCTTTTCTTTTAAATAA 78

DB 26 TTTCTTTTCTTTTAAATAA 9

RESULT 11

AAH69552

ID AAH69552 standard; cDNA; 311 BP.

XX AC AAH69552;

XX DT 19-SEP-2001 (first entry)

XX DE Human cervical cancer marker nucleic acid 826.

XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200142467-A2.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US33312.

XX PR 08-DEC-1999; 99US-0169881.

XX PR 21-DEC-1999; 99US-0171350.

XX PR 14-MAR-2000; 2000US-0189315.

XX PR 12-MAY-2000; 2000US-0203791.

XX PR 09-JUN-2000; 2000US-0210600.

XX PR 21-JUL-2000; 2000US-0220114.

XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Deeds J, Berger A, Zhao X;

XX DR WPI; 2001-375006/39.

XX New isolated nucleic acid for diagnosing and treating cervical cancer  
XX and for assessing and detecting compounds for treating the cancer -

XX PS Claim 1; Page 243; 1051pp; English.

XX The invention relates to novel genes (AAH68727-AAH73383) associated with

CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
CC polypeptides are useful: to assess if a patient is afflicted with  
CC cervical cancer or has a pre-malignant condition; to monitor the  
CC progression of cervical cancer or a pre-malignant condition in a patient;  
CC and to select and/or assess the efficacy of a compound or therapy for  
CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
CC useful for gene therapy.

XX Sequence 311 BP; 84 A; 67 C; 63 G; 94 T; 3 other;

Query Match 3.8%; Score 76; DB 22; Length 311;  
Best Local Similarity 100.0%; Pred. No. 2.7e-24;  
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1403 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCA 1462  
Db 24 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCA 83

QY 1463 GAGCTCAATAAGGTAC 1478

Db 84 GAGCTCAATAAGGTAC 99

#### RESULT 12

AAH70887

ID AAH70887 standard; cDNA; 313 BP.

XX AC AAH70887;

XX DT 19-SEP-2001 (first entry)

XX XX Human cervical cancer marker nucleic acid 2161.

XX XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX OS Homo sapiens.

XX XX WO200142467-A2.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US33312.

XX PR 08-DEC-1999; 99US-0169681.

XX PR 21-DEC-1999; 99US-0171350.

XX PR 14-MAR-2000; 2000US-0189315.

XX PR 12-MAY-2000; 2000US-0203791.

XX PR 09-JUN-2000; 2000US-0210600.

XX PR 21-JUL-2000; 2000US-0220114.

XX XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Deeds J, Berger A, Zhao X;

XX DR WPI; 2001-375006/39.

XX XX New isolated nucleic acid for diagnosing and treating cervical cancer

XX PT and for assessing and detecting compounds for treating the cancer .

XX PS Claim 1; Page 456; 1051pp; English.

XX XX The invention relates to novel genes (AAH68727-AAH73383) associated with

XX CC cervical cancer with cytostatic activity. The nucleic acids and encoded

XX CC polypeptides are useful: to assess if a patient is afflicted with

XX CC cervical cancer or has a pre-malignant condition; to monitor the

XX CC progression of cervical cancer or a pre-malignant condition in a patient;

XX CC and to select and/or assess the efficacy of a compound or therapy for

XX CC inhibiting cervical cancer in a patient. The nucleic acids may also be

XX CC useful for gene therapy.

XX XX Sequence 313 BP; 83 A; 70 C; 63 G; 97 T; 0 other;

Query Match

3.8%; Score 76; DB 22; Length 313;

Best Local Similarity 100.0%; Pred. No. 2.7e-24;  
Matches 76; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1403 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCA 1462  
Db 2 GTACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCA 61

QY 1463 GAGCTCAATAAGGTAC 1478

Db 62 GAGCTCAATAAGGTAC 77

#### RESULT 13

AAH72445

ID AAH72445 standard; cDNA; 299 BP.

XX AC AAH72445;

XX DT 19-SEP-2001 (first entry)

XX XX Human cervical cancer marker nucleic acid 3719.

XX XX Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.

XX OS Homo sapiens.

XX PN WO200142467-A2.

XX PD 14-JUN-2001.

XX PF 08-DEC-2000; 2000WO-US33312.

XX PR 08-DEC-1999; 99US-0169681.

XX PR 21-DEC-1999; 99US-0171350.

XX PR 14-MAR-2000; 2000US-0189315.

XX PR 12-MAY-2000; 2000US-0203791.

XX PR 09-JUN-2000; 2000US-0210600.

XX PR 21-JUL-2000; 2000US-0220114.

XX XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.

XX PI Schlegel R, Deeds J, Berger A, Zhao X;

XX DR WPI; 2001-375006/39.

XX XX New isolated nucleic acid for diagnosing and treating cervical cancer

XX PT and for assessing and detecting compounds for treating the cancer .

XX PS Claim 1; Page 696; 1051pp; English.

XX CC The invention relates to novel genes (AAH68727-AAH73383) associated with  
XX CC cervical cancer with cytostatic activity. The nucleic acids and encoded  
XX CC polypeptides are useful: to assess if a patient is afflicted with  
XX CC cervical cancer or has a pre-malignant condition; to monitor the  
XX CC progression of cervical cancer or a pre-malignant condition in a patient;  
XX CC and to select and/or assess the efficacy of a compound or therapy for  
XX CC inhibiting cervical cancer in a patient. The nucleic acids may also be  
XX CC useful for gene therapy.

XX SQ Sequence 299 BP; 82 A; 64 C; 57 G; 96 T; 0 other;

Query Match 3.7%; Score 74; DB 22; Length 299;

Best Local Similarity 100.0%; Pred. No. 2.2e-23;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1405 ACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCA 1464

Db 1 ACTCAACAGCAGCTGACTTATGTTTATTGGACATTGTGATACAGGAACGTGTTCCA 60

QY 1465 GCTCAATAAGGTAC 1478

Db .61 GCTCAATAAGGTAC 74

```
RESULT 14
AAH73007
XX AAH73007 standard; cDNA; 299 BP.
XX AC
XX AAH73007;
XX DT 19-SEP-2001 (first entry)
XX DE Human cervical cancer marker nucleic acid 4281.
XX KW Cervical cancer; cytostatic; pre-malignant condition; gene therapy; ss.
XX OS Homo sapiens.
XX PN WO200142467-A2.
XX PD 14-JUN-2001.
XX PF 08-DEC-2000; 2000WO-US33312.
XX PR 08-DEC-1999; 99US-0169681.
XX PR 21-DEC-1999; 99US-0171350.
XX PR 14-MAR-2000; 2000US-0189315.
XX PR 12-MAY-2000; 2000US-0203751.
XX PR 09-JUN-2000; 2000US-0210600.
XX PR 21-JUL-2000; 2000US-0220114.
XX PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX PI Schlegel R, Deeds J, Berger A, Zhao X;
XX WPI; 2001-375006/39.
XX DE New isolated nucleic acid for diagnosing and treating cervical cancer
XX PT and for assessing and detecting compounds for treating the cancer -
XX PS Claim 1; Page 941-942; 1051pp; English.
XX CC The invention relates to novel genes (AAH68727-AAH73383) associated with
XX CC cervical cancer with cytostatic activity. The nucleic acids and encoded
XX CC polypeptides are useful to assess if a patient is afflicted with
XX CC cervical cancer or has a pre-malignant condition; to monitor the
XX CC progression of cervical cancer or a premalignant condition in a patient;
XX CC and to select and/or assess the efficacy of a compound or therapy for
XX CC inhibiting cervical cancer in a patient. The nucleic acids may also be
XX CC useful for gene therapy.
XX SQ Sequence 299 BP; 82 A; 64 C; 56 G; 95 T; 2 other;
Query Match 3.7%; Score 74; DB 22; Length 299;
Best Local Similarity 100.0%; Pred. No. 2.2e-23;
Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1405 ACTCAACAGCAGCTACTTATGTTTATTTGACATTGATACAGGAAGCTTTCCAGA 1464
DB 1 ACTCAACAGCAGCTACTTATGTTTATTTGACATTGATACAGGAAGCTTTCCAGA 60
QY 1465 GCTCAATAAGGTAC 1478
DB 61 GCTCAATAAGGTAC 74
RESULT 15
ABN39970
ID ABN39970 standard; DNA; 60 BP.
XX AC
XX ABN39970;
XX DT 15-JUL-2002 (first entry)
XX DE Human spliced transcript detection oligonucleotide SEQ ID NO:12718.
XX KW Human; mouse; rat; splice transcript; detection; RNA transcript;
XX KW splice variant; transcriptome; oligonucleotide library; ss.
XX OS Homo sapiens.
XX PN WO200210449-A2.
XX PD 07-FEB-2002.
XX PF 20-JUL-2001; 2001WO-IB01903.
XX PR 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287724P.
XX PA (COMP-) COMPUGEN INC.
XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX WPI; 2002-257383/30.
XX DE New oligonucleotide libraries comprising oligonucleotides which
XX PT selectively hybridize to mRNAs transcribed from a transcription unit of
XX PT a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes -
XX PS Example 1; SEQ ID 12718; 47pp; English.
XX CC The present invention describes oligonucleotide libraries for detecting
XX CC messenger RNAs that populate a (sub-)transcriptome, where the
XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX CC several oligonucleotides, each capable of hybridising selectively to a
XX CC set of messenger RNAs transcribed from a given transcription unit of
XX CC the genome, which encodes one or more messenger RNA splice variants.
XX CC The oligonucleotide libraries are useful for detecting mRNAs from a
XX CC biological sample, in expression profiling studies, in qualitatively or
XX CC quantitatively characterising the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX CC transcriptomes. The libraries may also be used as specialised mini
XX CC libraries to detect transcripts of a sub-transcriptome under a
XX CC particular biological or pathological state, and so allowing the
XX CC detection of tissue- and pathology-specific genes such as those genes
XX CC only expressed in specific tissue under a specific pathological
XX CC condition; to detect developmental specific genes; and to detect RNA
XX CC transcripts and splice variants of a transcriptome of a patient suffering
XX CC from a particular disorder. ABN27253 to ABN59589 represent
XX CC oligonucleotide sequences from rats, humans and mice, which are used in
XX CC the exemplification of the present invention.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 60 BP; 22 A; 16 C; 9 G; 13 T; 0 other;
Query Match 3.0%; Score 60; DB 24; Length 60;
Best Local Similarity 100.0%; Pred. No. 5.5e-17;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 891 GTAGACCAGAACTGACCATTTCTCAGCCCTACTTCTCAGAAACAAACAGAGCTTTCAATG 950
DB 1 GTAGACCAGAACTGACCATTTCTCAGCCCTACTTCTCAGAAACAAACAGAGCTTTCAATG 60
Search completed: January 31, 2004, 16:17:08
Job time : 552 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: January 31, 2004, 06:31:47 ; Search time 86 Seconds

(without alignments)  
2386.548 Million cell updates/sec

Title: US-09-744-313A-1

Perfect score: 465

Sequence: 1 MYLIRCLIFRNTQKGESEF.....ELFELNKVQKVTSTSNM 465

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0

Xgapop 60.0 , Xgapext 60.0

Xgapop 6.0 , Xgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 569978 seqs, 220691566 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1135299

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DRV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09744313/runat\_30012004\_113613\_4617/app\_query.fasta\_1.647  
-DB=issued\_patents\_NA -SUFFIX=fastap -MINMATCH=0.1 -LOOCL=0  
-LIST=45 -DOCALIGN=200 -THR SCORE=quality -THR MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09744313 @CGN 1.1.56 @runat\_30012004\_113613\_4617 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOUERY -NEG SCORES=0 -WAIT -DSFLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : Issued Patents NA.\*

1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS.COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	98.1	1716	4	US-09-620-312D-922
2	307	66.0	1551	4	Sequence 922, App
3	8	1.7	85	2	Sequence 923, App
4	8	1.7	740	4	US-09-620-312D-923
5	8	1.7	966	4	Sequence 44, Appl
6	8	1.7	1239	4	Sequence 99, Appl
7	8	1.7	1842	4	US-09-342-681C-99
8	8	1.7	3043	4	Sequence 9495, Ap
9	8	1.7	152331	3	Sequence 9666, Ap
10	8	1.7	176373	3	Sequence 160, App
11	7	1.5	161	1	Sequence 14, Appl
12	7	1.5	161	1	Sequence 16, Appl
					Sequence 17, Appl
					Sequence 4, Appl
					Sequence 6, Appl

7	1.5	223	1	US-08-435-684A-43	Sequence 43, Appl
7	1.5	223	2	US-08-934-877A-43	Sequence 43, Appl
7	1.5	223	3	US-08-871-678C-43	Sequence 43, Appl
7	1.5	276	4	US-09-252-991A-991	Sequence 991, App
7	1.5	280	4	US-09-313-294A-998	Sequence 998, App
7	1.5	285	4	US-09-107-532A-3549	Sequence 3549, Ap
7	1.5	288	4	US-09-313-294A-3533	Sequence 3533, Ap
7	1.5	288	4	US-09-313-294A-5146	Sequence 5146, Ap
7	1.5	321	1	US-08-175-388-2	Sequence 2, Appl
7	1.5	321	2	US-08-779-620-2	Sequence 2, Appl
7	1.5	321	2	US-08-818-726-2	Sequence 2, Appl
7	1.5	321	3	US-09-018-584A-20	Sequence 20, Appl
7	1.5	365	3	US-09-326-039-1	Sequence 1, Appl
7	1.5	370	2	US-08-332-766A-8	Sequence 8, Appl
7	1.5	371	3	US-09-326-039-11	Sequence 11, Appl
7	1.5	421	2	US-08-332-766A-25	Sequence 25, Appl
7	1.5	475	4	US-09-221-017B-561	Sequence 561, App
7	1.5	597	2	US-08-332-766A-19	Sequence 19, Appl
7	1.5	627	3	US-09-328-111-194	Sequence 194, App
7	1.5	657	4	US-09-252-991A-13015	Sequence 13015, A
7	1.5	661	2	US-08-529-878B-37	Sequence 37, Appl
7	1.5	666	4	US-09-252-991A-11270	Sequence 11270, A
7	1.5	702	4	US-09-252-991A-6791	Sequence 6791, App
7	1.5	710	3	US-08-998-416-982	Sequence 982, App
7	1.5	725	3	US-08-998-416-984	Sequence 984, App
7	1.5	765	4	US-09-252-991A-6948	Sequence 6948, App
7	1.5	780	4	US-09-252-991A-6722	Sequence 6722, App
7	1.5	784	4	US-09-205-258-101	Sequence 101, App
7	1.5	797	3	US-08-981-803-13	Sequence 13, Appl
7	1.5	797	3	US-08-981-803-27	Sequence 27, Appl
7	1.5	797	3	US-08-983-440-13	Sequence 13, Appl
7	1.5	797	3	US-08-983-440-27	Sequence 27, Appl
7	1.5	797	4	US-09-367-895-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1

US-09-620-312D-922  
; Sequence 922, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, tian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunding  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662zel Nucleic Acids and  
; TITLE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt FL\_genes version 1.0  
; SEQ ID NO 922  
; LENGTH: 1716  
; TYPE: DNA

ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (71)..(1543)  
US-09-620-312D-922

## Alignment Scores:

Pred. No.: 0 Length: 1716  
Score: 456.00 Matches: 456  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.06% Indels: 0  
DB: 4 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-620-312D-922 (1-1716)

QY 10 PheArgAsnThrGlnLysArgGlyGluSerPheGlyIleSerArgIleGlySerIle 29  
DB 173 TTTCGGAACACACAGAAAAGGGAGAAATCATTTGGAATCAGCAGAAATAGGAGCAAAAT 232  
QY 30 LysGlyValPheLysSerThrThrMetGluGlyValaMetLeuProAsnTyrGlyValAla 49  
DB 233 AAGAGGATTTCAAAAGTACCAATGAGGAGGCTATGTTGCCCTAAATATGGTGTAGCT 292  
QY 50 GluGlyGluAspAspPheIleGluGlyIleValValMetGluAspAspSerProVal 69  
DB 293 GAAGGTGAAGATGATTTTATGAAGAAGGTATGTTGTAATGAAGATGATTTCTCCAGTG 352  
QY 70 GluAlaValSerThrProAsnThrProArgAsnLeuAlaIleThrPlysIleSerIlePro 89  
DB 353 GAGGCTGTGAGCACACCTAATCTCCCGAAACCTTGCTGCATGGAAATATAGCATTTCCA 412  
QY 90 TyrValAspPhePheGluAspProSerSerGluArgLysGluLysGluArgIlePro 109  
DB 413 TATGTAGACTTTTGTAGAGATCCCTCTCTGAAGAAGAGGAGAAAGAAAGAAATTCCT 472  
QY 110 ValPheCysIleAspValGluArgAsnAspArgArgAlaValGlyHisGluProGluHis 129  
DB 473 GTGTTTGTATGATGTTTGAAGAATATAGACAGGAGCTTGGACAGGCTGAACAT 532  
QY 130 TrpSerValTyrArgArgTyrLeuGluPheTyrValLeuGluSerLysLeuThrGluPhe 149  
DB 533 TGTCTGTCTATAGAAGATATCTGAATTTCTATGTTGTAATCAAACTAAACAGAAATTT 592  
QY 150 HisGlyAlaPheProAspAlaGlnLeuProSerLysArgIleIleGlyProLysAsnTyr 169  
DB 593 CATGGTGCAATTCCTGATGCCAGGCTCTCTTAAAGAGATCATTTGGCCCCCAAAATTTAT 652  
QY 170 GluPheLeuLysSerLysArgGluGluPheGlnGluTyrLeuGlnLysLeuGlnHis 189  
DB 653 GAATTTCTTAAAGTCAAGAGGAGAGATTCGAAGAATATCTACAGAAATCTTCTGAGCAT 712  
QY 190 ProGluLeuSerAsnSerGlnLeuLeuAlaAspPheLeuSerProAsnGlyGlyGluThr 209  
DB 713 CCAGAACTGAGTAATAGTCACTCTGGCAGACTTCTTCCCTTAATGGTGGGGAACA 772  
QY 210 GlnPheLeuAspLysIleLeuProAspValAsnLeuGlyLysIleLysSerValPro 229  
DB 773 CAATTTCTTGATAAGATATCAACAGATGTAATCTTGGGAAATATATAAATCTGTTCTCT 832  
QY 230 GlyLysLeuMetLysGluLysGlnHisLeuGluProPheIleMetAsnPheIleAsn 249  
DB 833 GGAAACTAATGAAGAGAGAGGTGAGCATTTTGGAACTTTTATCATGAAATTTTCATTAAT 892  
QY 250 SerCysGluSerProLysProLysProSerArgProGluLeuThrIleLeuSerProThr 269  
DB 893 TTTTGTGAGTCTCCAAAGCCTTAACCAAGTAGACAGAACTGACCAATTTCTCAGCCCTACT 952  
QY 270 SerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsnAlaAsnArgAlaGlu 289  
DB 953 TCAGAAACCAACAGAGGCTTTTCATGATCTGTTTAAATATATGCAACCGTGTCTGAA 1012  
QY 290 AsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMetThrValGluGlyVal 309

DB 1013 AATACAGAGAGAAAGCAAAATCAGAAATATTTATGAGGTGATGACTGTAGAGGAGTC 1072  
QY 310 TyrAspTyrLeuMetTyrValGlyArgValValPheGlnValProAspTrpLeuHisHis 329  
DB 1073 TATGATTAATCTGATGATGTAGGACGGTAGTTTCCAGGTTCTCTGACTGGCTTCAATAT 1132  
QY 330 LeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyrThrAspTyrTyr 349  
DB 1133 CTCTTAATGGAACTCGAATCTCTTTAAACACCCCTGAAATGTATACTGATTAATAT 1192  
QY 350 LeuGlnCysLysLeuGlnLeuPheGlnGluHisArgLeuValSerLeuLeuThrLeu 369  
DB 1193 CTTCACTGTAAACTAGAACAGCTATTTCCAGGAGCACCGTTTGTCTCACTCATTAACACTT 1252  
QY 370 LeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLysGlnLys 389  
DB 1253 CTGAGAGATGCTATATTTCTGTGAAACACACTGAACTCTCTCCAGATAGCAAAA 1312  
QY 390 GlyAlaLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuLeuValLysCys 409  
DB 1313 GGAGCAAAACAGACTTTTGAAGAAATGATGAATTAATTCAGATCTGTAGTCACTGT 1372  
QY 410 IleGlyGluGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuGlnGlnPro 429  
DB 1373 ATTTGTTGAAGAAACCAAGTATGAAGCATCAGACTTCTGTTGATGGCTTACAGCAACA 1432  
QY 430 ValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValIleGlnGluLeuPhePro 449  
DB 1433 GTACTCAACCAAGCAGCTGACTTATGTTTATTTGGAATTTGTGTATACAGGAACTGTTTCCA 1492  
QY 450 GluLeuAsnLysValGlnLysGluValThrSerValThrSerTropMet 465  
DB 1493 GAGCTCAATAGGTACAAAAGGAAGTTACCTCTGTGACATCTTGGATG 1540

## RESULT 2

US-09-620-312D-923  
; Sequence 923, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yunging  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE OF INVENTION: Polypeptides  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 923  
; LENGTH: 1551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:



QY 207 GAGTATTCAAAAGTACCAATGAGGAGCTATGTTGCTTAATTAATGTTAGTGTGAAG 266  
DB 237 GAGTATTCAAAAGTACCAATGAGGAGCTATGTTGCTTAATTAATGTTAGTGTGAAG 296  
QY 267 GTGAAGATGATTTTATGAAGAAGTATGTTGTTAATGAAGATGATTTCCAGTGGAGG 326  
DB 297 GTGAAGATGATTTTATGAAGAAGTATGTTGTTAATGAAGATGATTTCCAGTGGAGG 356  
QY 327 CTGTGAGCACACCTTAATCTCCCGAAACCTTCCTCATGCGAATTAATGCAATTCATATG 386  
DB 357 CTGTGAGCACACCTTAATCTCCCGAAACCTTCCTCATGCGAATTAATGCAATTCATATG 416  
QY 387 TAGACATTTTGTGAGATCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 446  
DB 417 TAGACATTTTGTGAGATCCCTCTCTGAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 476  
QY 447 TTTGTATTGATGTTGAAAGAAATGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 506  
DB 477 TTTGTATTGATGTTGAAAGAAATGATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 536  
QY 507 CTGTCTATGAGAGATCTTGAATCTTGTCTGATCTGATCTGATCTGATCTGATCTGATCT 566  
DB 537 CTGTCTATGAGAGATCTTGAATCTTGTCTGATCTGATCTGATCTGATCTGATCTGATCT 596  
QY 567 GTGCATTTCTGATGCCAGCTTCCTTCTAAGAGAGATCATTTGCCGCCCAAAATTAATGAT 626  
DB 597 GTGCATTTCTGATGCCAGCTTCCTTCTAAGAGAGATCATTTGCCGCCCAAAATTAATGAT 656  
QY 627 TCTTAAGTCAAAGAGGAGAGGTTCCAGAAATATCTACAGAAATCTTTCGAGCATCCAG 686  
DB 657 TCTTAAGTCAAAGAGGAGAGGTTCCAGAAATATCTACAGAAATCTTTCGAGCATCCAG 716  
QY 687 AACTGATTAATGATCTTCACTTCGAGAGCTTCTTCCCTTAATGTTGGGGGAGAGAGAG 746  
DB 717 AACTGATTAATGATCTTCACTTCGAGAGCTTCTTCCCTTAATGTTGGGGGAGAGAGAG 776  
QY 747 TCTTGTATGATGATCTTCACTTCGAGAGCTTCTTCCCTTAATGTTGGGGGAGAGAGAG 806  
DB 777 TCTTGTATGATGATCTTCACTTCGAGAGCTTCTTCCCTTAATGTTGGGGGAGAGAGAG 836  
QY 807 AACTTAATGAAGAGAGAGGTCAGCATTTGGAACTTTTATCATGAATTTTCAATATCTT 866  
DB 837 AACTTAATGAAGAGAGAGGTCAGCATTTGGAACTTTTATCATGAATTTTCAATATCTT 896  
QY 867 GTGAGTCTCAAAGCCTTAACCAAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 926  
DB 897 GTGAGTCTCAAAGCCTTAACCAAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 956  
QY 927 ABAACACAGAGAGCTTTTCAATGATCTGTTTAAATTAATGCAACCGTCTGAAATA 986  
DB 957 ABAACACAGAGAGCTTTTCAATGATCTGTTTAAATTAATGCAACCGTCTGAAATA 1016  
QY 987 CAGAGAGAAAGCAAAATCAGAAATTTTATGAGGAGTATGATCTGAGAGAGAGTCTATG 1046  
DB 1017 CAGAGAGAAAGCAAAATCAGAAATTTTATGAGGAGTATGATCTGAGAGAGAGTCTATG 1076  
QY 1047 ATTACCTGATGATGATGAGAGCGGATGTTTCCAGGTTCTCTGAGTGGCTTCTATCTCT 1106  
DB 1077 ATTACCTGATGATGATGAGAGCGGATGTTTCCAGGTTCTCTGAGTGGCTTCTATCTCT 1136  
QY 1107 TAATGGGAATCGAATCTCTTTAAACACCGCTGGAATGATATGATGATGATGATGATGAT 1166  
DB 1137 TAATGGGAATCGAATCTCTTTAAACACCGCTGGAATGATATGATGATGATGATGATGAT 1196  
QY 1167 AGTGAATCTAGAACAGCTTTTTCAGGAGACCGTTTGTCTCTCACTCAATCACTCTCA 1226  
DB 1197 AGTGAATCTAGAACAGCTTTTTCAGGAGACCGTTTGTCTCTCACTCAATCACTCTCA 1256  
QY 1227 GAGATGCTATTTCTGTGAG 1286  
DB 1257 GAGATGCTATTTCTGTGAG 1316  
QY 1287 CAAACAGAGCTTTTGAAGAAATGATGAAATTAATTCAGATCTGTTAGTCAAGTGTATG 1346

DB 1317 CAAAACAGAGCTTTTGAAGAAATGATGAAATTAATTCAGATCTGTTAGTCAAGTGTATG 1376  
QY 1347 GTGAAGAAACCAAGTATGAAGAGATCAGACTTCTGTTTGAAGGCTTACAGAAACCAAGTAC 1406  
DB 1377 GTGAAGAAACCAAGTATGAAGAGATCAGACTTCTGTTTGAAGGCTTACAGAAACCAAGTAC 1436  
QY 1407 TCAACAGAGAGCTGACCTTATGTTTATTTGACATTTGACATTTGACATTTGACATTTG 1466  
DB 1437 TCAACAGAGAGCTGACCTTATGTTTATTTGACATTTGACATTTGACATTTGACATTT 1496  
QY 1467 TCAATAGAGTACAAAAGGAAATTAACCTCTGTCATCTTCTGATGTAATGTAATGTAATG 1526  
DB 1497 TCAATAGAGTACAAAAGGAAATTAACCTCTGTCATCTTCTGATGTAATGTAATGTAATG 1556  
QY 1527 GTATAGAAATACCAATGAAATTTCTGCTGTCGAGAGGCTGTAAGAAATTAATCTTTTGG 1586  
DB 1557 GTATAGAAATACCAATGAAATTTCTGCTGTCGAGAGGCTGTAAGAAATTAATCTTTTGG 1616  
QY 1587 GTATATTTCTTATATATATATGATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1646  
DB 1617 GTATATTTCTTATATATATATGATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1676  
QY 1647 AATAAGAGCTAAACACAACTTAATGATTAATAA 1678  
DB 1677 AATAAGAGCTAAACACAACTTAATGATTAATAA 1708

## RESULT 2

US-09-620-312D-923  
; Sequence 923, Application US/09620312D  
; Patent No. 6569662  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Liu, Chenghua  
; APPLICANT: Asundi, Vinod  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Chen, Rui-hong  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Xue, Aidong J.  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Zhou, Ping  
; APPLICANT: Ma, Yungqing  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Wang, Zhiwei  
; APPLICANT: John Tillinghast  
; APPLICANT: Drmanac, Radoje T.  
; TITLE OF INVENTION: No. 6569662el Nucleic Acids and  
; FILE REFERENCE: 784CIP2B  
; CURRENT APPLICATION NUMBER: US/09/620,312D  
; CURRENT FILING DATE: 2000-07-19  
; PRIOR APPLICATION NUMBER: 09/552,317  
; PRIOR FILING DATE: 2000-04-25  
; PRIOR APPLICATION NUMBER: 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 1105  
; SOFTWARE: pt\_FL\_genes Version 1.0  
; SEQ ID NO 923\_  
; LENGTH: 1551  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURES:  
; NAME/KEY: CDS  
; LOCATION: (71)..(1378)  
US-09-620-312D-923

Query Match 46.1%; Score 918; DB 4; Length 1551;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 918; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-09-983-965-540  
; Sequence 540, Application US/09983965  
; Patent No. US20020137160A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Mengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 37-21(10297)C  
; CURRENT APPLICATION NUMBER: US/09/983,965  
; CURRENT FILING DATE: 2001-10-26  
; PRIOR FILING DATE: 1999-12-15  
; PRIOR FILING DATE: 1998-12-17  
; NUMBER OF SEQ ID NOS: 5912  
; SEQ ID NO 540  
; LENGTH: 402  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (339)  
; OTHER INFORMATION:  
; OTHER INFORMATION: Clone ID: 64-BOVMS1-017-Q1-E1-H12  
US-09-983-965-540

Alignment Scores:  
Pred. No.: 6.94e-73 Length: 402  
Score: 682.00 Matches: 127  
Percent Similarity: 99.25% Conservatives: 5  
Best Local Similarity: 95.49% Mismatches: 1  
Query Match: 27.99% Indels: 0  
DB: 9 Gaps: 0

US-09-744-313A-1 (1-465) x US-09-983-965-540 (1-402)

QY	205	AsnGlyGlyGluThrGlnPheLeuAspLysIleuProAspValAsnLeuGlyLysIle	224
DB	3	AATGGTGGGAAACACAGTCTTCTGAAGATACCTACAGACGTAAATCTTGGAAAT	62
QY	225	IleLysSerValProGlyLysLeuMetLysGlyLysGlnHisLeuGluProPheIle	244
DB	63	ATAAATCTGTTCTGGAAACTAATGAAGAGAAAGTCAACATTTGGACCTTTCATC	122
QY	245	MetAsnPheLeuAsnSerCysGluSerProLysProLysProSerArgProGluLeuThr	264
DB	123	ATGAATTCATTAATCTTGTGAATCTCCAAAGCCTAAACCGAGTAACCCAGAACTGACC	182
QY	265	IleLeuSerProThrSerGluAsnAsnLysLysLeuPheAsnAspLeuPheLysAsnAsn	284
DB	183	ATTCAGCCCTACTCTCAGAGATAATGAAGCTTTTAAATGATCTGTATAGAATAAT	242
QY	285	AlaAsnArgAlaGluAsnThrGluArgLysGlnAsnGlnAsnTyrPheMetGluValMet	304
DB	243	GCAACCGTCTGAGATACAGAAAGAGGCAAAATCAGAAATTAATTAATGGAATGATG	302
QY	305	ThrValGluGlyValTyrAspTyrLeuMetTyrValGlyArgValPheGlnValPro	324
DB	303	ACTGTAGAAGAGCTATGATTACTGATGATGTATGANGACAGGTGTTTCCAGATTCCT	352
QY	325	AspTrpLeuHisHisLeuLeuMetGlyThrArgIleLeu	337
DB	363	GACTGGCTTCATCATCTCTTAATGGGAACCTGAATCCTC	401

RESULT 3

US-10-085-783A-25415  
; Sequence 25415, Application US/10085783A  
; Publication No. US2004003784A1  
; GENERAL INFORMATION:  
; APPLICANT: ChondroGene Inc.

APPLICANT: Liew, C.C.  
; TITLE OF INVENTION: Compositions and Methods Relating to Osteoarthritis  
; FILE REFERENCE: 4231/2002  
; CURRENT APPLICATION NUMBER: US/10/085,783A  
; CURRENT FILING DATE: 2002-02-28  
; PRIOR APPLICATION NUMBER: US 60/305,340  
; PRIOR FILING DATE: 2001-07-13  
; PRIOR APPLICATION NUMBER: US 60/275,017  
; PRIOR FILING DATE: 2001-03-12  
; PRIOR APPLICATION NUMBER: US 60/271,955  
; PRIOR FILING DATE: 2001-02-28  
; NUMBER OF SEQ ID NOS: 58994  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 25415  
; LENGTH: 471  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (17)..(17)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (19)..(19)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (33)..(33)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (49)..(49)  
; OTHER INFORMATION: n is a, c, g, or t  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: (58)..(59)  
; OTHER INFORMATION: n is a, c, g, or t  
US-10-085-783A-25415

Alignment Scores:  
Pred. No.: 5.19e-69 Length: 471  
Score: 651.00 Matches: 131  
Percent Similarity: 94.24% Conservatives: 0  
Best Local Similarity: 94.24% Mismatches: 7  
Query Match: 26.71% Indels: 1  
DB: 12 Gaps: 0

US-09-744-313A-1 (1-465) x US-10-085-783A-25415 (1-471)

QY	328	HisHisLeuLeuMetGlyThrArgIleLeuPheLysAsnThrLeuGluMetTyr-ThrAs	347
DB	4	CATGACTGTAAANGAAGAACTCGAATCTTTTAAACACACCTGNAATGTATNTGA	63
QY	347	pTyrTyrLeuGlnCysLysLeuGluGlnLeuPheGlnGluHisArgLeuValSerLeu	367
DB	64	TTACTATCTTCAGTGTAAACTAGAACAGCTATTTCAGAGACCGTTTGGTCTCACTCAT	123
QY	367	ethrLeuLeuArgAspAlaIlePheCysGluAsnThrGluProArgSerLeuGlnAspLy	387
DB	124	AACACTTCTCAGAGATCTATATTCTGTGAAACACACTGAACCTCTCTCTCCAGATAA	183
QY	387	sGlnLysGlyValLysGlnThrPheGluGluMetMetAsnTyrIleProAspLeuVal	407
DB	184	GCAAAAGAGCAAAACAGACTTTTGAAGAAATGTAATTAATTCAGATCTGTAGT	243
QY	407	lLysCysIleGlyGluThrLysTyrGluSerIleArgLeuLeuPheAspGlyLeuG	427
DB	244	CAAGTGTATTGGTGAAGAAACCAAGTATGAAGCATCAGACCTCTGTGTGATGCTTACA	303
QY	427	nGlnProValLeuAsnLysGlnLeuThrTyrValLeuLeuAspIleValleGlnGluLe	447
DB	304	GCAACCAAGTACTCAACAGCAGCTGACTTATGTTTATTGGACATTTGTGATACAGGAAC	363